

Krepf, Matthias

## Wie analysieren ExpertInnen und NovizInnen Unterricht? Validierung des Konstrukts Analysekompetenz mittels Qualitativer Inhaltsanalyse

Bad Heilbrunn : Verlag Julius Klinkhardt 2019, 219 S. - (Klinkhardt forschung) - (Zugl.: Köln, Universität, Dissertation, 2018)



Quellenangabe/ Reference:

Krepf, Matthias: Wie analysieren ExpertInnen und NovizInnen Unterricht? Validierung des Konstrukts Analysekompetenz mittels Qualitativer Inhaltsanalyse. Bad Heilbrunn : Verlag Julius Klinkhardt 2019, 219 S. - (Klinkhardt forschung) - (Zugl.: Köln, Universität, Dissertation, 2018) - URN: urn:nbn:de:0111-pedocs-174571 - DOI: 10.25656/01:17457

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-174571>

<https://doi.org/10.25656/01:17457>

in Kooperation mit / in cooperation with:



<http://www.klinkhardt.de>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipt.de](mailto:pedocs@dipt.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Matthias Krepf

# **Wie analysieren ExpertInnen und NovizInnen Unterricht?**

**Validierung des Konstrukts Analysekompetenz  
mittels Qualitativer Inhaltsanalyse**

Krepf

# Wie analysieren ExpertInnen und NovizInnen Unterricht?



Matthias Krepf

# Wie analysieren ExpertInnen und NovizInnen Unterricht?

Validierung des Konstrukts Analysekompetenz  
mittels Qualitativer Inhaltsanalyse

Verlag Julius Klinkhardt  
Bad Heilbrunn • 2019

**k**

*Meinen Eltern, Barbara und Ulrich, gewidmet.*

Die vorliegende Arbeit wurde im Januar 2018 von der Humanwissenschaftlichen Fakultät der Universität zu Köln als Dissertation angenommen.

Gutachter: Prof. Dr. Wilfried Plöger, Prof. Dr. Petra Herzmann.

Tag der Disputation: 24.01.2018

Dieser Titel wurde in das Programm des Verlages mittels eines Peer-Review-Verfahrens aufgenommen.  
Für weitere Informationen siehe [www.klinkhardt.de](http://www.klinkhardt.de).

Bibliografische Information der Deutschen Nationalbibliothek  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation  
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten  
sind im Internet abrufbar über <http://dnb.d-nb.de>.

2019.kg. © by Julius Klinkhardt.

Das Werk ist einschließlich aller seiner Teile urheberrechtlich geschützt.

Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung  
des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen,  
Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Druck und Bindung: AZ Druck und Datentechnik, Kempten.

Printed in Germany 2019.

Gedruckt auf chlorfrei gebleichtem alterungsbeständigem Papier.

ISBN 978-3-7815-2312-8

# Danksagung

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Januar 2018 angenommen.

Zum Abschluss des Promotionsverfahrens wird meist ein letzter, abschließender Textbaustein der Arbeit hinzugefügt: die Danksagung, denn am Ende einer langen, herausfordernden und intensiven Reise lohnt sich ein Blick zurück, um den Personen zu danken, die maßgeblich an der Entstehung der Arbeit beteiligt waren.

Zuallererst gilt mein Dank meinem Doktorvater Herrn Prof. Dr. Wilfried Plöger für die intensive, motivierende Unterstützung in allen Phasen des Promotionsprozesses. Danken möchte ich ihm für seinen engagierten Einsatz, seine Geduld, die stets offene Tür für Fragen, seine konstruktiven Rückmeldungen und die gemeinsamen Gespräche, die ich als Ermutigung und Motivation verstanden habe. Ich danke ihm von Herzen für die hervorragende Betreuung und die lehrreiche Zeit an seinem Lehrstuhl!

Des Weiteren möchte ich meinem Kollegen Herrn Prof. Dr. Daniel Scholl für die erfolgreiche Zusammenarbeit danken und dass er mir mit Rat und Tat zur Seite stand.

Ferner möchte ich Alina Diederichs, Eric Berndt und Max Peltzer herzlich danken. Sie haben mir, im Rahmen ihrer Tätigkeit als studentische Hilfskräfte, tatkräftig bei der Transkription und Kodierung der Interviews geholfen.

Einen besonderen Dank möchte ich Herrn PD Dr. Rainer Wisbert für unsere zahlreichen kontroversen, anregungsreichen Debatten sowie motivierenden Gespräche aussprechen.

Darüber hinaus möchte ich mich bei meiner Familie und meinen Freunden bedanken. Ein aufrichtiges Dankeschön dafür, dass sie mich stets unterstützt haben sowie mit Geduld und Verständnis meinen Weg verfolgten.

Mein abschließender Dank gilt meinen Eltern, Barbara und Ulrich. Ich danke ihnen für ihre Liebe und ihre fortwährende, bedingungslose Unterstützung, die mir die Kraft gab, dieses Projekt in Angriff zu nehmen und erfolgreich abzuschließen. Aus diesem Grund möchte ich ihnen dieses Buch widmen!





# Kurzzusammenfassung

Die vorliegende Arbeit fokussiert auf die Analysekompetenz von Lehrpersonen, definiert als jene Fähigkeit, „die Qualität ... von beobachtetem Unterricht im Hinblick auf dessen Lernwirksamkeit angemessen erfassen und bewerten zu können“ (Plöger & Scholl, 2014a, S. 88). Im Sinne dieser Definition stellen konkrete Analyseleistungen empirische Manifestationen des theoretisch postulierten Konstrukts Analysekompetenz dar.

Zwischen den Annahmen eines solchen Konstrukts und den jeweils erhobenen Daten klafft – methodologisch gesehen – allerdings eine prinzipiell unüberwindbare Lücke, weil Daten empirischer, Konstrukte dagegen rein gedanklicher bzw. theoretischer Natur sind. Aufgrund dieses prinzipiellen Unterschiedes ist man im praktischen Forschungsprozess permanent mit der Frage der Validität konfrontiert: Repräsentieren die Daten in adäquater Weise das, was das theoretische Konstrukt (hier Analysekompetenz) postuliert?

Eine Variante der evidenzbasierten Validitätsprüfung stellt der *Mixed-Methods-Ansatz* dar, dessen Potenzial ich in meiner Untersuchung genutzt habe. Ihren Ausgangspunkt bildeten zwei Validierungsbelege, die von Plöger et al. (2015) und Plöger und Scholl (2014a) durch Konfirmatorische Faktorenanalysen erbracht wurden und bei denen sich zeigte: (1) Analysekompetenz kann als ein zweidimensionales Konstrukt interpretiert werden; dabei repräsentiert die inhaltliche Dimension das verfügbare pädagogische und fachliche Wissen, während die formale Dimension für die Komplexität der Informationsverarbeitung steht. (2) Im Sinne des sogenannten *well-known-group-Vergleichs* (s. z.B. Cronbach & Meehl, 1955) konnte zudem nachgewiesen werden, dass sich Personen, denen man vorab eine unterschiedliche hohe Analysekompetenz unterstellen kann, auch tatsächlich in den erzielten Testscores bei der Messung der Analysekompetenz unterscheiden.

Beide Validierungsbelege stellten wichtige Orientierungspunkte für meine Untersuchung dar, indem ich ebenfalls das von Plöger und Scholl definierte Verständnis von Analysekompetenz zugrunde legte und die verwendete Stichprobe neun ExpertInnen und neun NovizInnen umfasste, um starke Unterschiede in den Analyseleistungen sichtbar machen zu können. Als Stimulus wurde den ProbandInnen eine videographierte Unterrichtsstunde vorgeführt. Auf Grundlage dieser Unterrichtsstunde wurden anschließend *fokussierte, leitfadengestützte* Interviews geführt.

Die Auswertung der Interviews erfolgte in zwei Schritten: (1) Die transkribierten Interviews wurden im Sinne der Qualitativen Inhaltsanalyse nach Mayring anhand eines umfangreichen Kategoriensystems ausgewertet. (2) In einem weiteren Schritt wurden die ermittelten Aktivierungen pädagogischen und fachlichen Wissens weiter quantifiziert, um Signifikanzprüfungen, Reliabilitätsanalysen sowie Trennschärfe- und Korrelationsberechnungen durchführen zu können.

Die Ergebnisse meiner Untersuchung lieferten eindeutige Belege, mit denen die zweidimensionale Struktur des Konstrukts Analysekompetenz auch auf qualitativem Wege mittels der hier angewandten Qualitativen Inhaltsanalyse repliziert werden konnte. Darüber hinaus konnten die im Sinne des *well-known-group-Vergleichs* vermuteten Unterschiede zwischen ExpertInnen und NovizInnen bei der Analyse von Unterricht ebenfalls repliziert werden.

# Abstract

The present work focuses on the analytical competence of teachers, defined as the ability of teachers to properly perceive and assess their own and others' lessons in terms of the effectiveness of learning. In the sense of this definition, concrete analytical performances represent empirical manifestations of the theoretically postulated construct analytical competence. From a methodological point of view, however, there is in principle an insurmountable gap between the assumptions of such a construct and the data collected, because data are empirical, while constructs are of pure conceptual or theoretical nature. Due to this difference in principle, the question of validity arises permanently in the practical research process: do the data adequately represent what the theoretical construct (here analytical competence) claims?

One variant of evidence-based validity checking is the *mixed-methods approach*, being used in my research project presented in this book. The starting point were two proofs of validity, provided by Plöger et al. (2015) and Plöger and Scholl (2014a) by confirmatory factor analyses showing that: (1) analytical competence can be defined as a construct having two dimensions – a content and a formal dimension. The content dimension is made up of two factors: pedagogical knowledge and content knowledge. The formal dimension of the construct is defined using three main level and five substages of complexity with which teachers can process information relevant for teaching. (2) In the sense of the so-called *well-known-group comparison* (e.g. Cronbach & Meehl, 1955), it could also be proven that persons, assumed to have a different level of analytical competence in advance, actually did so in the test scores obtained by measuring the analytical competence. Both successful proofs of validity provided important reference points for my study, by sharing the concept of analytical competence as defined by Plöger and Scholl, and by using a sample which comprised nine experts and nine novices in order to make visible large differences in the analytical performances. As a stimulus the participants were shown a physics lesson on optics in which the Snell's law of refraction was being studied. In the subsequent interview, subjects were asked to comment on the lesson they had observed.

The analysis of the data was undertaken in two steps: (1) First, the interviews were transcribed in full and analysed using the qualitative content analysis method (Mayring, 2014). (2) In the second stage, the ascertained activations of pedagogical and content knowledge were quantitatively processed in order to calculate the reliability of the codings and the correlations of the categories between the factors of the content and formal dimensions. Furthermore, I used the Mann-Whitney Test to check whether the experts and novices differed significantly in their analytical performance.

The results of my investigation provided clear evidence to replicate the two-dimensional structure of the construct analytical competence in a qualitative way using the content analysis applied here. In addition, the well-known-group comparison confirmed the supposed differences between experts and novices in the analysis of lessons.

# Inhalt

<b>Einleitung .....</b>	<b>11</b>
<b>1 Theoretische Begründung des Konstrukts Analysekompetenz .....</b>	<b>19</b>
1.1 Lehrerkompetenzforschung – Wissen als zentrale Voraussetzung für gute Analyseleistungen .....	19
1.2 Definition der Facetten des didaktisch-methodischen Wissens .....	23
1.3 Definition der Facetten des fachlichen (incl. fachdidaktischen) Wissens .....	27
1.4 Das Paradigma der Expertiseforschung – Komplexität der Informationsverarbeitung .....	32
1.4.1 Ergebnisse der Expertiseforschung .....	32
1.4.2 Fünfstufiges Kompetenzmodell nach Plöger und Scholl .....	35
1.5 Methode und Ergebnisse der Untersuchung von Plöger und Scholl (2014) .....	38
1.5.1 Anlage der Studie .....	38
1.5.2 Zentrale Ergebnisse der Studie von Plöger und Scholl (2014) .....	40
1.6 Quantitativer Validierungsversuch von Plöger, Scholl und Seifert (2015) .....	41
1.6.1 Modellprüfung mittels Konfirmatorischer Faktorenanalysen .....	41
1.6.2 Über CFA ermittelte faktorielle Struktur von Analysekompetenz .....	46
1.6.3 Relevanz der Ergebnisse für den durchzuführenden qualitativen Validierungsversuch .....	48
<b>2 Methodisches Vorgehen .....</b>	<b>49</b>
2.1 Qualitative Inhaltsanalyse als methodischer Zugriff .....	49
2.2 Vorüberlegungen zur Datengenerierung .....	52
2.3 Datengenerierung mittels Interviews .....	53
2.3.1 Das Interview als Datengenerierungsquelle .....	53
2.3.2 Fokussierte Interviews .....	54
2.3.3 Der Interviewleitfaden .....	55
2.3.4 Auswahl der TeilnehmerInnen .....	57
2.3.5 Durchführung der Interviews .....	59
2.4 Datenaufbereitung – Transkription .....	60
2.5 Kategorienbasierte Kodierung der Transkripte .....	62
2.5.1 Entwicklung eines Kategoriensystems .....	62
2.5.2 Bestimmung der Analyseeinheiten .....	63
2.5.3 Konsensuelles Kodieren zur Überprüfung der Bildung und Anwendung von Kategorien .....	64
2.5.4 Prüfung der Reliabilität .....	66
2.5.5 Prüfung der Validität .....	69
2.6 Ablaufmodell der Qualitativen Inhaltsanalyse .....	70

<b>3 Theoretische Herleitung der Kategorien .....</b>	<b>73</b>
3.1 Kategorien für didaktisch-methodisches Wissen .....	75
3.1.1 Bezüge zum Modell der Struktur- und Prozessqualität von Unterricht ....	75
3.1.2 Prozessmerkmale von Unterricht .....	77
3.1.3 Kategorien für didaktisch-methodisches Wissen .....	85
3.2 Kategorien für fachliches (incl. fachdidaktisches) Wissen .....	93
3.2.1 Das Vorgehen zur Bildung von Kategorien für fachliches und fachdidaktisches Wissen .....	93
3.2.2 Gegenstand und Methode der Physik als indirekte Bezugsgröße der Kategorienbildung .....	95
3.2.3 Auswahl der Kategorien für fachliches und fachdidaktisches Wissen ...	100
3.2.4 Kategorien für fachliches (incl. fachdidaktisches) Wissen .....	109
3.3 Kategorien für die formale Qualität des Wissens (Verstehenstiefe) .....	120
3.3.1 Das Kompetenzmodell von Plöger und Scholl als Bezugsrahmen .....	120
3.3.2 Bestimmung der Kategorien für die formale Qualität des Wissens (Verstehenstiefe) in Entsprechung zu den fünf Stufen des Kompetenzmodells von Plöger und Scholl .....	123
3.3.3 Kategorien und Ankerbeispiele für die formale Qualität des Wissens ...	126
<b>4 Hypothesen .....</b>	<b>131</b>
<b>5 Ergebnisdarstellung .....</b>	<b>135</b>
5.1 Überprüfung der Inter- und Intracoderreliabilität .....	135
5.2 Ergebnisse der Qualitativen Inhaltsanalyse .....	136
5.2.1 Unterschiede zwischen ExpertInnen und NovizInnen .....	137
5.2.2 Unterschiede zwischen ExpertInnen und NovizInnen hinsichtlich des fachlichen Hintergrundes .....	140
5.3 Vorstellung beispielhafter Aussagen .....	146
5.3.1 Beispielhafte Aussagen für die inhaltliche Dimension .....	147
5.3.2 Beispielhafte Aussagen für die formale Dimension .....	153
5.4 Quantifizierung der Ergebnisse .....	156
5.4.1 Reliabilitätsanalysen der bisherigen Ergebnisse .....	158
5.4.2 Endgültige Ergebnisse der Qualitativen Inhaltsanalyse .....	164
5.4.3 Beurteilung der Unterschiede zwischen ExpertInnen und NovizInnen ...	170
<b>6 Zusammenfassung .....</b>	<b>173</b>
<b>7 Ausblick .....</b>	<b>179</b>
7.1 Das Potenzial des Mixed-Methods-Ansatzes zur Konstruktvalidierung .....	179
7.2 Ein Ausblick auf die „Amalgamthese“ Shulmans .....	181
<b>Verzeichnisse</b>	
Literaturverzeichnis .....	185
Abbildungsverzeichnis .....	209
Tabellenverzeichnis .....	209
<b>Anhang .....</b>	<b>213</b>

# Einleitung

Die vorliegende Arbeit berichtet über eine qualitative Studie, die ich als Wissenschaftlicher Mitarbeiter im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Forschungsprojekts „Analysekompetenz – Diagnose und Entwicklung“ (Leitung: Prof. Dr. Wilfried Plöger & Dr. Daniel Scholl; in Kooperation mit Dr. Andreas Seifert) durchgeführt habe. Unter Analysekompetenz verstehen Plöger und Scholl die Fähigkeit, „die Qualität ... von beobachtetem Unterricht im Hinblick auf dessen Lernwirksamkeit angemessen erfassen und bewerten zu können“ (Plöger & Scholl, 2014a, S. 88).

Paradigmatisch gesehen, ist dieses DFG-Projekt dem *Experten-Novizen-Paradigma* verpflichtet (s. z.B. Bromme, 1997; Bromme & Rheinberg, 2006; Gruber, 2001; Gruber & Stöger, 2011; Krauss, 2011; Krauss & Bruckmaier, 2014; König, 2010; Mulder & Gruber, 2011). Das gilt auch für meine Studie, in der ich Interviews mit ExpertInnen und NovizInnen geführt und mittels Qualitativer Inhaltsanalyse (s. z.B. Mayring, 2014, 2015) ausgewertet habe, wobei meine zentrale Forschungsfrage in Anlehnung an die Definition des Begriffs der Analysekompetenz von Plöger und Scholl lautete: Welche Dispositionen ermöglichen es ExpertInnen im Vergleich zu NovizInnen, Unterricht angemessen unter dem Aspekt der Lernwirksamkeit zu analysieren?

Der Ablauf der einzelnen Phasen meiner Studie lässt sich im Groben folgendermaßen skizzieren: Zu Beginn der Studie wurde den TeilnehmerInnen (9 ExpertInnen/9 NovizInnen) eine Physik-Stunde (Thema: Snelliussches Brechungsgesetz) präsentiert. Die im Anschluss an diese Videopräsentation geführten Interviews wurden transkribiert und in einem weiteren Schritt anhand von Kategorien analysiert. Diese Kategorien repräsentierten die kognitiven Dispositionen der ProbandInnen in Form ihres *pädagogischen* und *fachlichen* Wissens, sodass sich die konkrete Fragestellung ergab: Welche Facetten pädagogischen und fachlichen Wissens ermöglichen ExpertInnen und NovizInnen eine angemessene Analyse von Unterricht?

In systematischer Hinsicht ist die meiner Arbeit zugrunde liegende Forschungsfrage somit in der *Lehrerkompetenzforschung* angesiedelt (s. z.B. Baumert & Kunter, 2006; Herzmann & König, 2016; König, 2010, 2016; Mulder & Gruber, 2011; Terhart, 2007). Wenn man aus der Perspektive dieser Forschung mit Baumert und Kunter den Kern der Tätigkeit von Lehrpersonen im Unterrichten sieht (s. Baumert & Kunter, 2011a, 2013), und diese Tätigkeit permanent im Kreislauf der Planung, Durchführung und Analyse von Unterricht erfolgreich gemeistert werden muss, dann stellt die Analysekompetenz in diesem Zyklus eine zentrale Kompetenz dar, weil sie der Lehrperson „Aufschluss über die Qualität ihrer Planung und die Effektivität des durchgeführten Unterrichts“ (Plöger & Scholl, 2016, S. 181) gibt.

In methodologischer Hinsicht stellen Kompetenzen *Konstrukte* dar: Sichtbare Leistungen in Form beobachtbarer Daten (hier Analyseleistungen) werden auf latente Fähigkeiten (kognitive, motivationale Dispositionen) zurückgeführt. Ob die beobachteten Daten als Folge solcher latenten Dispositionen gedeutet werden dürfen, bedarf einer differenzierten Rechtfertigung, die letztlich auf die Kernfrage hinausläuft: Können die erhobenen Daten

mit den theoretischen Annahmen interpretiert werden, die das jeweilige Konstrukt definieren? Im Falle meiner Studie sollte also geprüft werden, ob die durch eine kategoriengeleitete Analyse der Interviews erzeugten Daten Belege für die theoretische Grundannahme darstellen, dass das Gelingen der Analyse von Unterricht vom verfügbaren pädagogischen und fachlichen Wissen der ProbandInnen abhängt.

Diese Frage zielt methodologisch gesehen auf die Frage nach der *Validität des Konstrukts Analysekompetenz*. Validität stellt, neben der Objektivität und Reliabilität, das bedeutendste und komplexeste Gütekriterium empirischer Sozialforschung dar (s. z.B. Blömeke, 2013; Hartig, Frey & Jude, 2012; Wilhelm & Kunina-Habenicht, 2015).<sup>1</sup>

Im traditionellen Verständnis wurde und wird Validität als die Eigenschaft eines Tests aufgefasst, der dann als valide gilt, wenn er das misst, was er zu messen vorgibt bzw. „misst, was er messen soll“ (s. z.B. Eid & Schmidt, 2014; Friedrichs, 1990; Lienert & Raatz, 1998; Rost, 2004; Sedlmeier & Renkewitz, 2013).

In den vergangenen Jahren hat sich allerdings die Bedeutung von Validität geändert, so dass die ursprüngliche Auffassung von Validität (als Eigenschaft eines Tests) inzwischen abgelehnt wird: „It is incorrect to use the unqualified phrase, the validity of the test“ (AERA, APA, & NCME, 2014, S. 11). Infolgedessen wird Validität nicht mehr als Testeigenschaft aufgefasst, sondern Validität „refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.“ (ebd., S. 11) Das bedeutet, dass nicht mehr das Testverfahren selbst überprüft wird, sondern die *Interpretationen* von Testwerten *validiert* werden (s. Frey, 2015; Furr & Bacharach, 2008; Kane, 2013). *Validierung* kann, in Anlehnung an Frey (2015), somit als „Sammlung relevanter Informationen“ verstanden werden, „die belastbare, wissenschaftlich abgesicherte Unterstützung angestrebter Testwertinterpretationen liefern“.<sup>2</sup>

Einen entscheidenden Anteil an diesem veränderten Verständnis von Validität haben die Arbeiten von Messick (1989, 1990, 1995). Nach Messick ist Validität „an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.“ (Messick, 1989, S. 13; 1995, S. 741) Validität wird somit zum allgemeinen Bewertungsmaßstab erklärt, um Interpretationen von Testwerten zu beurteilen.

Auch im Sinne dieses neueren Verständnisses von Validität behalten „traditionelle“ Arten der Validitätsprüfung allerdings ihre Funktion. Das gilt insbesondere für diese drei Arten der Validitätsprüfung (s. AERA, APA, & NCME, 1985; Blömeke, 2013; Döring & Bortz, 2016; Bühner, 2011; Hartig et al., 2012):

- 
- 1 Die Objektivität umfasst die *Unabhängigkeit* der Messung bzw. der Testergebnisse von den Rahmenbedingungen der Testung (s. z.B. Blömeke, 2013, Bühner, 2011). Mit der Reliabilität wird die *Zuverlässigkeit* der Messung bzw. der Ergebnisse erfasst. Während eine hohe Reliabilität eine objektive Erhebung erfordert, ist die Reliabilität wiederum eine Voraussetzung für eine hohe Validität (s. z.B. Bühner, 2011; Furr & Bacharach, 2008). Die Validität nimmt somit eine Sonderstellung ein, denn „Objektivität und Reliabilität sind nur notwendige Minimalanforderungen an ein Messinstrument. Das Hauptziel ist dagegen die Konstruktion möglichst valider Messinstrumente“ (Dickmann, 2009, S. 256).
  - 2 Das Zitat stammt aus einem Vortrag von A. Frey aus dem Jahre 2015 zur 80. Tagung der *AEPF* (Arbeitsgruppe für Empirische Pädagogische Forschung) in Göttingen. Ich danke ihm für die Bereitstellung des Foliensatzes.

- *Inhaltsvalidität*: Dabei wird „das Verhältnis zwischen dem zu erfassenden Merkmal und den Test- bzw. Iteminhalten“ (Hartig et al., 2012, S. 146) durch *ExpertInnenurteile* geprüft.
- *Kriteriumsvalidität*: Hierbei wird das Ergebnis eines zu validierenden Konstrukts mit einem externen Kriterium (Außenkriterium) in Beziehung gesetzt (s. Balderjahn, 2003; Blömeke, 2013).<sup>3</sup>
- *Konstruktvalidität*: Der Grundgedanke ist, dass „Testergebnisse vor dem Hintergrund eines theoretischen Konstrukts“ (Hartig et al., 2012, S. 153) interpretiert werden. Es wird also geprüft, ob die „Gültigkeit der Interpretation empirischer Daten als Indikatoren für theoretische Konstrukte“ (Döring & Bortz, 2016, S. 98) vorliegen.<sup>4</sup>

Während die Konstruktvalidierung Mitte des 20. Jahrhunderts zunächst als Ergänzung zur Inhalts- und Kriteriumsvalidität gesehen wurde, stellt sie seit den 1970er Jahren einen „übergeordneten Zugang zur Bestimmung der Validität“ (Hartig et al., 2012, S. 153; Blömeke, 2013; Cronbach & Meehl, 1955; Frey, 2015) dar: „Ein Ziel bei der Untersuchung der Konstruktvalidität besteht vereinfacht ausgedrückt in der Überprüfung theoretischer Annahmen über Zusammenhangsstrukturen latenter Konstrukte anhand empirischer Daten.“<sup>5</sup> (Hartig et al., 2012, S. 156) Eine solche Überprüfung der Konstruktvalidität kann aber nicht hinreichend im Rahmen einer einzelnen empirischen Studie beantwortet werden. Vielmehr gilt: „A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation“ (AERA et al., 2014, S. 21) von Daten für die spezifische Verwendung.

In dieser Forderung, verschiedene Arten der Evidenzprüfung zu nutzen, sehe ich eine große Chance zur Beantwortung meiner Forschungsfrage und gehe deshalb mit Leech und Dellinger davon aus, „that construct validity and the construct validation process are relevant to mixed methods research“ (Dellinger & Leech, 2007, S. 321).

Grundlegendes Ziel einer Mixed-Methods-Untersuchung ist es, dass „the researcher might put quantitative and qualitative findings side-by-side to check how congruent they are“ (Bryman, 2008, S. 163). Dabei „geht es vor allem darum, eine pragmatische Verknüpfung von qualitativer und quantitativer Forschung zu ermöglichen“ (Flick, 2011b, S. 76, s. auch Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998, 2003b).<sup>6</sup>

3 Externe Kriterien können dabei *Ergebniskriterien* (z.B. Schulnoten), *Verhaltenskriterien* (z.B. Feedbackverhalten von Lehrpersonen) oder *Eigenschaftskriterien* (z.B. motivationale Einstellungen von Lehrpersonen) sein (s. z.B. Blömeke, 2013).

4 Eine bedeutende Arbeit in diesem Zusammenhang haben Cronbach und Meehl (1955) vorgelegt. Unter einem Konstrukt verstehen sie Folgendes: „A construct is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting a test is a construct.“ (Cronbach & Meehl, 1955, S. 283).

5 Weitere Möglichkeiten zur Überprüfung der Konstruktvalidität könnten, neben der Prüfung der faktoriellen Struktur eines Konstrukts, die „Prüfung der Konstruktrepräsentation über die Vorhersage von Itemschwierigkeiten und die Prüfung der Verortung des Konstrukts in einem nomologischen Netzwerk“ (Blömeke, 2013, S. 7) sein.

6 Die Entwicklung des Mixed-Methods-Ansatzes kann nur im Hinblick auf die historische Entwicklung in der empirischen Sozialforschung erklärt werden (s. Kelle, 2008, 2014). Die dort seit den 1920er Jahren stattfindende Entwicklung hat zu zwei unabhängigen Methodentraditionen (qualitativ/quantitativ) geführt, die sich „im Rahmen eigener methodologischer Traditionen und Denkschulen entwickelt“ (Kelle, 2014, S. 154) haben. Doch die damit verbundenen unterschiedlichen Forschungsziele und Gütekriterien führten zu teilweise



Eine einheitliche Definition von Mixed-Methods liegt jedoch bisher nicht vor (s. Johnson, Onwuegbuzie, & Turner, 2007). Es existiert aber ein grundlegend geteiltes Verständnis, was man unter Mixed-Methods-Ansätzen verstehen kann: „Mixed methods research is the type of research in which a researcher or team of researchers combine elements of qualitative and quantitative research approaches (e.g. use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purpose of breadth and depth of understanding and corroboration ...” (Johnson et al., S. 123).

Im Vergleich zu seinem zentralen Stellenwert „the concept of validity has been addressed sparingly in the mixed methods literature” (Dellinger & Leech, 2007, S. 314) seit längerer Zeit. Mittlerweile liegen allerdings sowohl methodologische Konzepte von Mixed-Methods-Ansätzen zur Validitätsprüfung (z.B. Dellinger & Leech, 2007; Leech, Dellinger, Brannagan, & Tanaka, 2010; Onwuegbuzie, Bustamante, & Nelson, 2010) als auch empirische Studien vor, in denen Mixed-Methods-Ansätze zur Beurteilung der Validität eingesetzt wird (s. z.B. Daigneault & Jacob, 2014; Latcheva, 2009; Luyt, 2011; Morell & Tan, 2009). Trotzdem wird immer noch konstatiert, dass „more contributions about why and how to mix methods for validation purposes are needed” (Daigneault & Jacob, 2014, S. 6). In dieser Arbeit nutze ich das Potenzial des Mixed-Methods-Ansatzes, indem ich empirische Daten auf *qualitativem* Wege sammle und prüfe, ob sie die Validität des Konstrukts Analysekompetenz stützen. Die Ergebnisse dieses qualitativen Vorgehens setzte ich dann in Beziehung zu einem *quantitativen* Validierungsversuch, der von Plöger, Scholl und Seifert (2015) durchgeführt wurde, um über den Vergleich der Ergebnisse festzustellen, ob beide Ansätze, der qualitative und der quantitative, die zugrunde liegenden theoretischen Annahmen des Konstrukts Analysekompetenz valide stützen.

Beide Studien sind in einem *sequential explanatory design* (Ivankova, Creswell & Stick, 2006) miteinander verknüpft.<sup>7</sup> Bei einem sequentiellen Design sind grundsätzlich zwei Varianten denkbar: Entweder es wird mit einer qualitativen Studie oder es wird mit einer quantitativen Studie begonnen. Die Wahl des Beginns bestimmt, ob man von einem „qualitativ-vertiefenden Design“ oder einem „quantitativ-verallgemeinernden Design“ sprechen kann (Kuckartz, 2014, S. 77). Im vorliegenden Fall wurde mit dem quantitativen Validierungsversuch begonnen (s. Plöger, Scholl & Seifert, 2015), an den sich die hier vorliegende qualitativ-vertiefende Untersuchung anschließt.

Die Frage der Konstruktvalidierung steht in diesem Fall in engem Zusammenhang mit dem Konzept der *Kompetenz* (s. z.B. Gruber, 2004; Klieme & Leutner, 2006; Oser & Renold, 2005; Shavelson, 2013; Terhart, 2007; Zlatkin-Troitschanskaia & Seidel, 2011). Dieses Konzept lässt sich mit Blömeke, Gustafsson und Shavelson (2015) durch zwei Pole charakterisieren:

---

kontrovers geführten Debatten, die in ihrer „Hoch-Zeit“ mit dem Slogan „paradigm wars“ (Gage, 1989) umschrieben wurden. Diesem „paradigm war“ stellt sich die Mixed-Methods-Bewegung entgegen und erklärt ihn für beendet (Flick, 2011b, S. 76). Seit den 1990er Jahren zielt diese Bewegung darauf, eine „third methodological movement“ (Tashakkori & Teddlie, 2003a) zu etablieren.

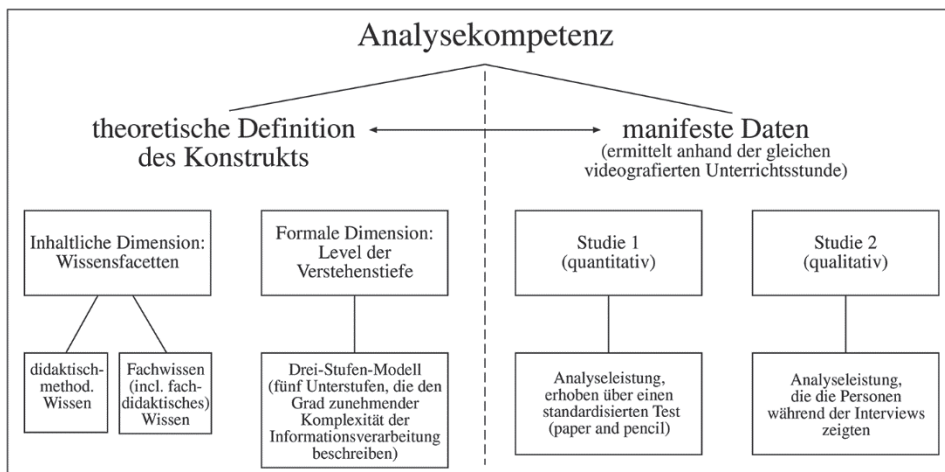
- 7 Bisher liegt keine einheitliche Systematik vor, wie Mixed-Methods-Studien durchgeführt werden sollten, sondern es konkurrieren eine Vielzahl unterschiedlicher Taxonomien miteinander (s. z.B. Kelle, 2014; Kuckartz, 2014; Morse, 1991; Schreier & Echterhoff, 2013). Einen Systematisierungsvorschlag haben unter anderem Creswell und Plano Clark (2011) unternommen. Sie identifizieren sechs zentrale Untersuchungsanlagen für Mixed-Methods-Designs, von denen das explanatory (sequential) design eines ist. Weiterhin unterscheiden sie zwischen einem *convergent*, einem *exploratory*, einem *embedded*, einem *transformative* und einem *multiphase design*.



Den einen Pol bilden *latente Dispositionen* (kognitive und affektiv-motivationale Einstellungen), den anderen *manifeste Leistungen* (beobachtbares Verhalten), die als Indikatoren für die latenten Dispositionen betrachtet werden können.<sup>8</sup>

Prinzipiell klappt eine unüberwindbare Lücke zwischen diesen beiden Polen: Latente Dispositionen sind gedanklicher bzw. theoretischer und manifeste Leistungen dagegen empirischer Natur. Infolge dieses grundlegenden Unterschieds ist die Frage der Validität im praktischen Forschungsprozess permanent virulent: Repräsentieren die Daten in adäquater Weise das, was die theoretische Definition des Konstrukts postuliert?

Zur Veranschaulichung des Zusammenhangs zwischen diesem Verständnis von Kompetenz, dem realisierten sequential explanatory design und dem oben beschriebenen Verständnis des Konstrukts Analysekompetenz zu veranschaulichen, greife ich auf die folgende Grafik zurück:



**Abb. 1:** Definition des Konstrukts Analysekompetenz und zeitliche Abfolge der beiden Validierungsstudien

<sup>8</sup> Blömeke et al. (2015) unterscheiden in diesem Kontinuum drei Ebenen von Kompetenzkonstrukten: disposition; situation specific skills und performance. Dabei beziehen sich *dispositions* auf affektive, motivationale und kognitive Konstrukte. Diese Konstrukte werden meistens mittels Tests (Wissen) oder Persönlichkeits-tests (affektive und motivationale Einstellungen/Überzeugungen) in Studien untersucht. *Situation specific skills* sind dahingegen mehr aktionsorientiert, und die Art der kognitiven Prozesse ist in diesem Bereich deutlich komplexer, da Prozesse wie Wahrnehmung, Interpretation und Entscheidungsfindung untersucht werden. Infolge der Komplexität dieser situationsspezifischen Fähigkeiten wird auch die Messung komplizierter. In vielen Fällen wird ein Video als Impulsgeber eingesetzt, um diese Fähigkeiten zu testen. In der vorliegenden Arbeit wird das Konstrukt Analysekompetenz als situation specific skill interpretiert. Als dritten Bereich benennen Blömeke et al. (2015) die *performance*. Diese ist zwar (mehr oder weniger) direkt beobachtbar, aber auch hier liegt ein latentes Konstrukt zugrunde, das definiert, was es bedeutet, ob eine Person in der Lage ist, die Anforderungen in einem bestimmten Bereich zu meistern (hier: die Fähigkeit, Unterricht angemessen analysieren zu können).

Diese drei Ebenen bauen hierarchisch aufeinander auf: Dispositions sind Voraussetzungen für situation-specific skills, welche wiederum Voraussetzungen für die performance sind. Somit wird die Struktur des Wissens (dispositions) in der Struktur der Analysekompetenz (situation specific skill) gespiegelt.

Auf der rechten Seite der Grafik ist die Abfolge der beiden Studien dokumentiert, durch die die entsprechenden manifesten Daten erhoben wurden. Auf der linken Seite sind die theoretischen Annahmen skizziert, durch die Analysekompetenz als zweidimensionales Konstrukt definiert wird. Dabei repräsentiert die erste Dimension die *inhaltliche* Qualität des Wissens, die zwei Wissensfacetten als Faktoren umfasst: das didaktisch-methodische Wissen einerseits und das *fachliche* (incl. *fachdidaktische*) Wissen andererseits. Die zweite Dimension stellt die *formale* Qualität des Wissens dar, die den Grad der für die Analyseleistungen notwendigen Komplexität der Informationsverarbeitung abbildet und sich in drei Faktoren (Analytische Kompetenz, Synthetische Kompetenz und Prozesskompetenz) mit insgesamt fünf Substufen unterteilen lässt. Im Sinne dieses Verständnisses von Analysekompetenz sind nur solche Personen in der Lage, Unterricht angemessen zu analysieren, die sowohl in *inhaltlicher* als auch in *formaler* Hinsicht über eine hinreichend hohe Qualität des Wissens verfügen.

Diese theoretischen Annahmen wurden von Plöger et al. (2015) auf quantitativem Wege mittels Konfirmatorischer Faktorenanalysen (*confirmatory factor analysis* – CFA) geprüft (Studie 1). Die dadurch ausgewiesene und mit entsprechenden Fit-Werten belegte faktorielle Struktur des Konstrukts Analysekompetenz dient nun in meiner Studie als hypothetische Basis, an das das erkenntnisleitende Interesse meiner Studie anknüpft (Studie 2). Mittels Qualitativer Inhaltsanalyse soll geprüft werden, ob sich die über CFA ermittelte Faktorenstruktur auch auf qualitativem Wege replizieren lässt. Inhaltlich gesehen orientieren sich demnach beide Verfahren (Qualitative Inhaltsanalyse, CFA) an einem *identischen* Verständnis von Analysekompetenz. In methodischer Hinsicht stellen sie dagegen im Sinne eines Mixed-Methods-Designs zwei *unterschiedliche* Wege zur Validierung dieses Konstrukts dar.

## Aufbau der Arbeit

Im Folgenden skizziere ich den Aufbau der vorliegenden Arbeit, die sich insgesamt in sieben Kapitel gliedert.

In Kapitel 1 werde ich – ausführlicher als in dieser Einleitung – den theoretischen und empirischen Hintergrund meiner Untersuchung darstellen, sodass dieses Kapitel inhaltlich zweigeteilt ist: Im ersten Teil werden die theoretischen Annahmen, die dem Konstrukt Analysekompetenz zugrunde liegen, präsentiert. Daran anschließend werden in einem zweiten Teil die empirischen Studien von Plöger und Scholl (2014a, 2014b) und Plöger et al. (2015) vorgestellt, um die Relevanz der Ergebnisse dieser Untersuchungen für meine Arbeit zu erörtern.

Kapitel 2 stellt den methodischen Zugriff der Arbeit dar, der sich in Kürze so zusammenfassen lässt: Auf Grundlage einer videographierten Unterrichtsstunde wurden *fokussierte, leitfadengestützte* Interviews geführt, per Tonbandgerät aufgezeichnet und anschließend *vollständig* transkribiert. Um die Interviews auswerten und die faktorielle Struktur des Konstrukts Analysekompetenz abbilden zu können, habe ich mich für die Qualitative Inhaltsanalyse (s. z.B. Mayring, 2010, 2014, 2015) entschieden. Mit ihr werden große Textmengen durch Anwendung von Kategorien auf zentrale Aussagen reduziert. Die Qualitative Inhaltsanalyse ist also insofern eine äquivalente Methode zur CFA, weil sie letztlich ein *datenreduzierendes* Verfahren darstellt.

Im Anschluss an den methodischen Teil werde ich in Kapitel 3 die Kategorien meiner Untersuchung theoretisch herleiten. Dazu ist es notwendig, auf die Forschung zur professionellen Handlungskompetenz von Lehrpersonen einzugehen und einen Überblick über aktuelle Konzeptualisierungen der professionellen Kompetenz von Lehrpersonen zu geben, wie sie z.B. in Studien wie *COACTIV (Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz;* Blum, Krauss, & Neubrand, 2011; Krauss et al., 2004; Kunter, Klusmann, & Baumert, 2009; Kunter et al., 2011), *TEDS-M (Teacher Education and Development Study: Learning to Teach Mathematics*, Blömeke, Kaiser, & Lehmann, 2010a, 2010b), *LEK (Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden*, König & Seifert, 2012) und *MT 21 (Mathematics Teaching in the 21st Century*, Blömeke, Kaiser, & Lehmann, 2008) vorgenommen wurden. Die einzelnen Teilkapitel sind an der faktoriellen Struktur des Konstrukts Analysekompetenz orientiert. In den einzelnen Teilkapiteln werde ich Bezüge zur empirischen Lehr-Lern-Forschung, zur Allgemeinen Didaktik und zur Kognitionpsychologie herstellen, um die Bildung meiner Kategorien sowohl empirisch als auch theoretisch zu begründen. Zum Abschluss der jeweiligen Teilkapitel werde ich die Kategorien vorstellen, die letztlich das „Herzstück“ der Untersuchung darstellen. Die Arbeit ist als Validierungsversuch geplant. Validierungen können aber nur vorgenommen werden, wenn entsprechende Hypothesen vorliegen, die geprüft werden können. Daher werde ich in Kapitel 4 die der Arbeit zugrunde liegenden Hypothesen vorstellen.

In Kapitel 5 werden die Untersuchungsergebnisse präsentiert. Dieses Kapitel untergliedert sich in einen qualitativen und einen quantitativen Teil. Im qualitativen Teil werde ich entsprechende Textbelege liefern, die die Annahmen meiner Arbeit stützen. Diese qualitativen Daten werden anschließend *quantitativ* weiterverarbeitet. Dabei wird jedes einzelne Kodierereignis als Indiz für die spezifische Wissensaktivierung einer Wissensfacette ver-

standen, sodass jedes Kodierereignis eine richtige, von den KodiererInnen akzeptierte Anwendung von Wissen repräsentiert. Unterschiede, die sich bei den Zahlen bzw. Summen ergeben, werden dann als entsprechende Unterschiede hinsichtlich der Ausprägung der Wissensaktivierung und somit der Analyseleistung ausgelegt.

In Kapitel 6 werden die zentralen Ergebnisse entlang der Fragestellung zusammengefasst und vor dem Hintergrund der Hypothesen meiner Untersuchung beurteilt.

Zum Schluss der Arbeit gebe ich in Kapitel 7 einen kurzen Ausblick auf weiterführende Forschungsfragen, die sich aus der theoretischen und forschungsmethodologischen Anlage der Untersuchung ergeben könnten.

# **1. Theoretische Begründung des Konstrukts Analysekompetenz**

In der vorangegangenen Einleitung habe ich das übergreifende Ziel meines Forschungsvorhabens umrissen, das auf die Beantwortung der Frage abzielt, ob sich die manifesten Daten, die ich mit meiner qualitativen Studie erhoben habe, als valide Belege für die faktorielle Struktur des Konstrukts Analysekompetenz interpretieren lassen.

Sollte dies der Fall sein, würde die von mir durchgeführte Qualitative Inhaltsanalyse den quantitativen Validierungsversuch von Plöger et al. (2015) stützen und damit exemplarisch das Potenzial von Mixed-Methods-Ansätzen für Fragen der (Konstrukt-)Validierung untermauern. Mit beiden methodischen Ansätzen wären dann empirische Daten auf unterschiedlichen Wegen generiert und ausgewertet worden, die kompatibel mit den theoretischen Annahmen des Konstrukts Analysekompetenz sind.

Um die zentrale Frage meines Forschungsvorhabens beantworten zu können, sind zwei aufeinander abgestimmte Schritte zu gehen: In einem ersten Schritt werden in diesem Kapitel die theoretischen Annahmen, die das Konstrukt Analysekompetenz charakterisieren, expliziert und vor dem Hintergrund der gegenwärtigen Forschungssituation legitimiert. Anschließend wird in einem zweiten Schritt (s. Kap. 2) dann zu begründen sein, inwiefern die Qualitative Inhaltsanalyse als geeignetes methodisches Instrument erscheint, adäquate empirische Daten für die Beantwortung der Fragestellung zu erzeugen und zu prüfen, ob diese manifesten Daten das theoretische Konstrukt decken.

In diesem Kapitel geht es zunächst um den ersten Schritt der theoretischen Legitimation des Konstrukts Analysekompetenz und damit um die Vorstellung der von Plöger und Scholl (2014a) sowie Plöger et al. (2015) vorgenommenen theoretischen Modellierung des Konstrukts Analysekompetenz. Anschließend werden das von ihnen gewählte Verfahren der Datenerhebung und -auswertung und die Ergebnisse ihres quantitativen Validierungsversuchs in Kürze dargestellt bzw. erläutert.

## **1.1 Lehrerkompetenzforschung – Wissen als zentrale Voraussetzung für gute Analyseleistungen**

In der gegenwärtigen Forschung zur Kompetenz von Lehrpersonen wird Wissen als eine unabdingbare Disposition zur professionellen Bewältigung täglich anfallender Aufgaben und Herausforderungen verstanden (Baumert & Kunter, 2011a, 2013; König, 2010; Shulman, 1986, 1987). Das gilt nicht nur für die Planung und Durchführung von Unterricht, sondern auch für dessen Analyse (s. z.B. Blomberg, Renkl, Sherin, Borko, & Seidel, 2013; Borko, Jacobs, Eiteljorg, & Pittman, 2008; Brophy, 2004; Rosaen, Carlisle, Mihocko, Melnick, & Johnson, 2013). Denn kognitionspsychologisch gesehen lässt sich die Analyse

von Unterricht als ein Prozess der Aufnahme und Verarbeitung von Informationen verstehen, der *wissensbasiert* und *wissensgesteuert* abläuft. Demnach hängt die Ausprägung der Analysekompetenz vom Umfang und von der Organisation des verfügbaren Wissens ab. Um diese These theoretisch zu begründen und das Konstrukt Analysekompetenz zu legitimieren, greife ich in diesem Abschnitt zunächst auf die gegenwärtige Diskussion um die Bedeutung des Faktors Wissen zurück, um dann im folgenden Schritt zu erläutern, welche spezifischen Wissensfacetten für die Entwicklung und Ausprägung der Analysekompetenz eine wichtige Rolle spielen.

Die Diskussion um die Bedeutung des *Professionswissens* von Lehrpersonen beschäftigt ForscherInnen seit mehreren Jahrzehnten. Den Ausgangspunkt dieser Diskussion stellen die Klassifikationsvorschläge von Shulman (1986, 1987) dar, die zunächst im angloamerikanischen Raum intensiv rezipiert wurden. Daran schloss sich – auch im deutschsprachigen Raum – eine ausführliche und langanhaltende Diskussion zur Beschreibung der notwendigen Wissensbasis von Lehrpersonen an. Stellvertretend seien hier die Arbeiten von Bromme (1992, 2008) und der Forschungsgruppe um Weinert (Weinert, Schrader, & Helmke, 1990) genannt, die den (damaligen) wissenschaftlichen Diskurs prägten.<sup>9</sup>

Letztlich und trotz aller Disparatheit hinsichtlich der Konzeptualisierungen des professionellen Wissens von Lehrkräften hat diese Diskussion dazu geführt, dass *allgemeines pädagogisches Wissen*, *Fachwissen* und *fachdidaktisches Wissen* als unentbehrliche Aspekte professionellen Lehrerhandelns gesehen werden (s. z.B. Baumert & Kunter, 2011a, 2013; Blömeke, Suhl, & Kaiser, 2011; Borko, 2004; König, 2010; König et al., 2014; Munby, Russel, & Martin, 2002; Voss, Kunina-Habenicht, Hoehne, & Kunter, 2015; Voss, Kunter, Seiz, Hoehne, & Baumert, 2014) und dass diese Aspekte deshalb auch in entsprechenden Forschungsprojekten eine zentrale Rolle einnehmen. Das gilt z.B. für Projekte wie *COACTIV*, *TEDS-M*, *LEK* und *MT 21*.

Wie diese drei Aspekte – allgemeines pädagogisches Wissen, Fachwissen und fachdidaktisches Wissen – näher präzisiert werden, sei exemplarisch am Projekt *COACTIV* gezeigt. Das Anliegen von *COACTIV* besteht darin, „die individuellen Merkmale zu identifizieren, die Lehrkräfte für die erfolgreiche Bewältigung ihrer beruflichen Aufgaben benötigen“ (Baumert & Kunter, 2011a, S. 29), um darauf aufbauend ein Modell professioneller Kompetenz von Lehrpersonen zu entwickeln.

Zur Identifizierung entsprechender Merkmale und zur Entwicklung des Modells werden verschiedene (theoretische und empirische) Ansätze aufgegriffen, von denen zwei von zentraler Bedeutung sind: Erstens wird ein *professionsspezifischer* Zugang gewählt, der auf den oben angegebenen Arbeiten von Shulman (und Bromme) basiert. Zweitens wird

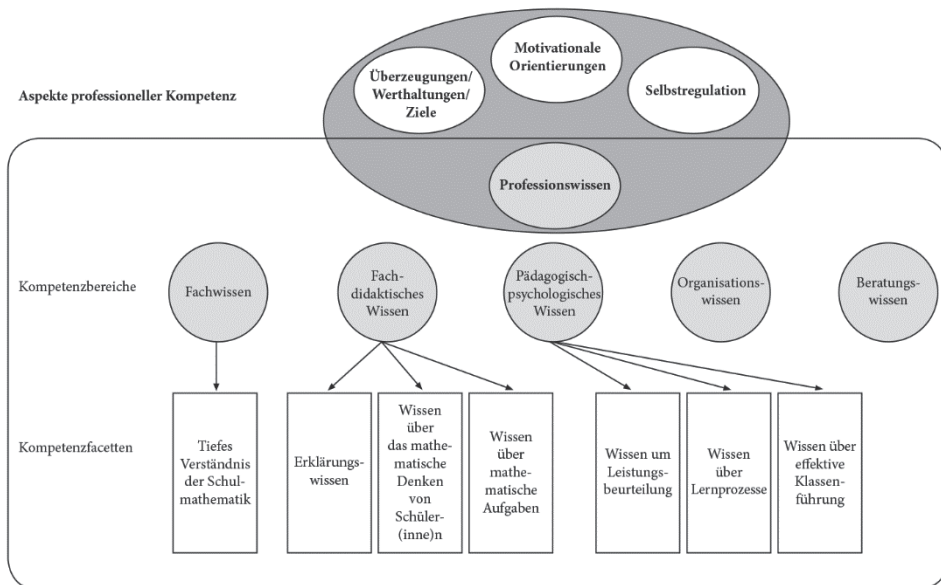
9 In den 1980er Jahren unterschied Shulman (1986) zunächst zwischen *general pedagogical knowledge* (allgemeines, fächerübergreifendes pädagogische Wissen), *subject matter content knowledge* (Fachwissen), *pedagogical content knowledge* (fachdidaktisches Wissen) und *curricular knowledge* (curriculares Wissen). Später wurde diese Klassifikation durch weitere Wissensfacetten erweitert, sodass diese nun zusätzlich *knowledge of learners and their characteristics* (Vorwissen der SchülerInnen), *knowledge of educational contexts* (Wissen über den Bildungskontext), sowie *knowledge of educational ends, purposes, and values, and their philosophical and historical grounds* (Wissen über pädagogische Ziele, Zwecke) umfasste (s. Shulman, 1987, S. 8). Bromme (1992, 1997) griff die Klassifikationsvorschläge auf, unterschied aber nur zwischen fünf inhaltlichen Kategorien: *allgemeines pädagogisches Wissen*; *Fachwissen*; *curriculares Wissen*; *Philosophie des Schulfaches*; *fachspezifisch-pädagogisches Wissen*. Weinert et al. (1990) beschränkten sich dagegen auf vier Aspekte des Wissens, über die Lehrpersonen verfügen sollten: *Klassenführungswissen*, *unterrichtsmethodisches Wissen*, *Fachwissen* sowie *diagnostisches Wissen*.

zur Beschreibung der „Bewältigung beruflicher Anforderungen“ (Baumert & Kunter, 2011a, S. 31) der *weite* Kompetenzbegriff von Weinert verwendet, der neben kognitiven Aspekten zusätzlich noch motivationale, metakognitive und selbstregulative Merkmale professionellen Handelns berücksichtigt:

The theoretical construct of action competence comprehensively combines those intellectual abilities, content-specific knowledge, cognitive skills, domain-specific strategies, routines and subroutines, motivational tendencies, volitional control systems, personal value orientations, and social behaviors into a complex system. Together, this system specifies the prerequisites required to fulfill the demands of a particular professional position.

(Weinert, 2001, S. 51)

Im Anschluss an diese Definition Weinerts wurde im COACTIV-Projekt ein Kompetenzmodell entwickelt (s. Abb. 2), in dem Baumert und Kunter vier *Kompetenzaspekte* unterscheiden, die zugleich die oberste Ebene des Modells bilden: (1) Überzeugungen/Werthaltungen/Ziele, (2) motivationale Orientierungen, (3) Selbstregulation und (4) das Professionswissen.



**Abb. 2:** Das Kompetenzmodell von COACTIV mit Spezifikation für das Professionswissen (aus: Baumert & Kunter, 2011a, S. 32)

Dieser vierte Aspekt, das Professionswissen, wird in fünf weitere Kompetenzbereiche unterteilt: Fachwissen, fachdidaktisches Wissen, pädagogisch-psychologisches Wissen, Organisationswissen und Beratungswissen. Die ersten drei Kompetenzbereiche werden dann in weitere Kompetenzfacetten untergliedert:



- *Fachwissen* wird dabei als tiefes Verständnis der Schulmathematik verstanden.
- *Fachdidaktisches Wissen* wird unterteilt in Erklärungswissen, Wissen über das mathematische Denken von Schülerinnen und Schülern und das Wissen über mathematische Aufgaben.
- *Pädagogisch-psychologisches Wissen* wird aufgefasst als Wissen über Leistungsbeurteilung, Lernprozesse sowie effektive Klassenführung.

An dieser Konzeptualisierung von Wissen orientiert sich auch das DFG-Projekt „Analysekompetenz – Diagnose und Entwicklung“ von Plöger und Scholl (2014a), und zwar in Verbindung mit der Annahme, dass die Ausprägung der Analysekompetenz von Lehrpersonen durch derartige Wissensfacetten bedingt ist. Auch die von mir durchgeführte Qualitative Inhaltsanalyse steht in der Spur dieser Annahme, wobei ich zugleich die von Plöger und Scholl aus forschungspragmatischen Gründen vollzogenen folgenden *Einschränkungen* und *Modifikationen* teile:

- Plöger und Scholl beschränken sich auf drei Facetten des Professionswissens (pädagogisch-psychologisches Wissen, Fachwissen und fachdidaktisches Wissen). Die Begriffe *Organisationswissen* und *Beratungswissen* werden *nicht berücksichtigt*, da sie für die Analyse von Unterricht keine unmittelbare Relevanz besitzen.
- Zudem wird der Begriff des pädagogisch-psychologischen Wissens bei Plöger und Scholl durch die Bezeichnung *didaktisch-methodisches Wissen* ersetzt, um den *handlungsnäheren* Charakter dieses Wissens stärker zu betonen (s. Plöger & Scholl, 2013, 2014a). Denn bei der Analyse von Unterricht stehen Situationen und Handlungen im Fokus, deren Lernwirksamkeit beurteilt werden muss.
- Darüber hinaus nehmen sie keine Differenzierung zwischen fachlichem und fachdidaktischem Wissen vor, sondern subsumieren unter dem Etikett „Fachwissen“ zugleich fachdidaktisches Wissen. Dafür sprechen insbesondere zwei Gründe: Zum einen hat die Diskussion gezeigt, dass es sich um schwer separierbare Wissensaspekte handelt, die mittels Paper-Pencil-Tests und hinreichend differenzierender Aufgabentypen zwar getrennt erhoben werden können, die aber umso eher „verschmolzen“ vorliegen (s. Shulman, 1987), je näher man an das Lehrerhandeln kommt. Denn das *Handeln* von Lehrpersonen ist nicht direkt an den jeweiligen disziplinären Wissensstrukturen orientiert, sondern erfolgt prinzipiell *kontext-* und *situationsgebunden* (s. Bromme, 1992). Zum anderen konnten im Rahmen der oben angesprochenen Forschungsprogramme auf empirischem Wege hohe Korrelationen zwischen dem fachlichen und dem fachdidaktischen Wissen nachgewiesen werden. Beispielsweise liegt bei COACTIV eine Korrelation von .79 für die Gesamtstichprobe und .91 für Gymnasiallehrkräfte vor (Krauss et al., 2011, S. 148f.). In TEDS-M beträgt sie .70 (Blömeke, Kaiser, Döhrmann, & Lehmann, 2010, S. 212), in MT21 wurde eine Korrelation von .81 (Blömeke, Seeber et al., 2008, S. 69) ermittelt. Als Konsequenz derartig hoher Zusammenhänge zwischen Fachwissen und fachdidaktischem Wissen werden in neueren Kompetenzmodellierungen beide Wissensbereiche als eine *Einheit* zusammengefasst (s. für den Bereich der Mathematikdidaktik z.B. Lindmeier, 2011, S. 105).



## 1.2 Definition der Facetten des didaktisch-methodischen Wissens

Didaktisch-methodisches Wissen ist *generisches* fächerübergreifendes Wissen, über das alle Lehrpersonen neben ihrem spezifischen fachlichen und fachdidaktischen Wissen verfügen sollten, um Unterricht professionell planen, durchführen und analysieren zu können. In der Planungsphase dient dieses didaktisch-methodische Wissen dazu, den (möglichen) Unterrichtsverlauf strukturell vorwegzunehmen, sodass er dann in der unterrichtlichen Kommunikation und Interaktion mit den SchülerInnen unter dem Gesichtspunkt der Lernwirksamkeit umgesetzt werden kann. Der Weg von der Planung bis hin zur Durchführung von Unterricht ist also dadurch bestimmt, dass künftiges Handeln *wissensgesteuert* bzw. *wissensbasiert* entworfen wird.

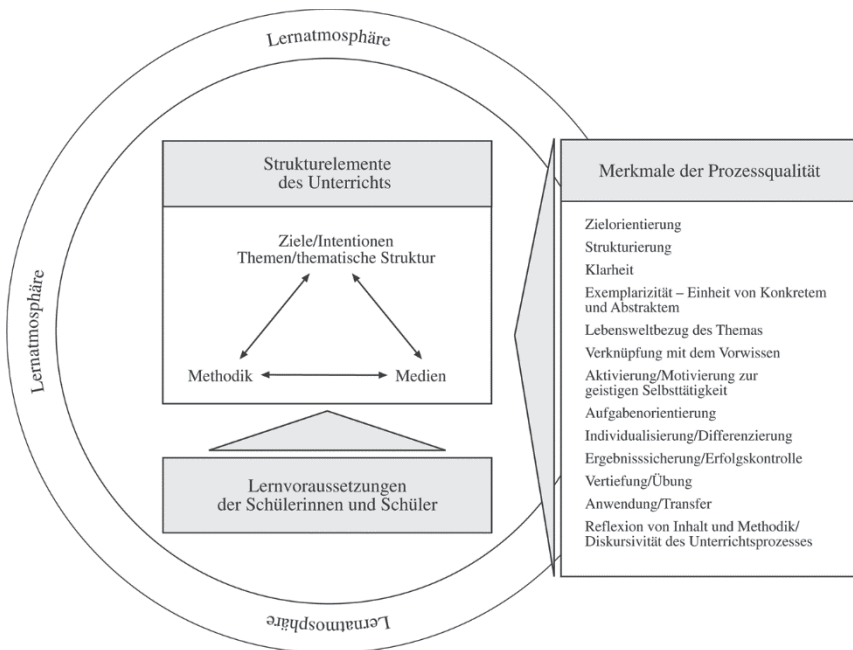
Im Rahmen der nachträglichen Analyse gleicht die Lehrperson ihr eigenes Handeln mit den Lernergebnissen der Schülerinnen und Schüler ab. Während also Handeln in der Planungsphase von Unterricht wissensbasiert *erzeugt* wird, muss dieses Handeln im Rahmen der nachträglichen *Analyse* wissensbasiert *rekonstruiert* werden. In diesem Prozess der Analyse wird dann das didaktisch-methodische (und das fachliche) Wissen zur Beurteilung der Lernwirksamkeit des vorangegangenen Unterrichts genutzt (s. Plöger & Scholl, 2013, 2014b).<sup>10</sup>

Das für die Planung und Analyse notwendige didaktisch-methodische Wissen wird von Plöger und Scholl im Rahmen eines von ihnen so bezeichneten *Modells der Struktur- und Prozessqualität* von Unterricht konzeptualisiert. In der folgenden Abbildung sind auf der linken Seite die Elemente der *Strukturqualität* und auf der rechten Seite die Merkmale der *Prozessqualität* von Unterricht abgebildet. Gerahmt wird das Modell durch den Faktor der

<sup>10</sup> Unter einem lernwirksamen Unterricht wird hier ein Unterricht verstanden, der das sinnstiftende Lernen seitens der SchülerInnen fördert. Dabei stellt das sinnstiftende Lernen einen mentalen Konstruktionsprozess dar, bei dem zwischen (spezifischen) Wissenselementen neue Verknüpfungen hergestellt werden (s. Reynolds, 1992), wodurch neue zusammenhängende kognitive Strukturen entstehen (s. Collins, Greeno, & Resnick, 2001). In diesem Lernprozess sollten die neu zu generierenden Verbindungen inhaltlich spezifisch gerichtet sein, damit die SchülerInnen die benötigten Wissenselemente aus ihrem Vorwissen abrufen, die neu aufzunehmenden Informationen damit abgleichen und somit relevante von irrelevanten Informationen trennen können (s. Mayer, 2002; Reynolds, 1992). Die Lernenden sollen also nicht willkürliche Beziehungen stiften, sondern solche, „die strukturell im Hinblick auf die Bildung der neuen kognitiven Struktur notwendig sind“ (Plöger & Scholl, 2015, S. 4). Durch das Erkennen der spezifischen Beziehungen erhalten sie somit Einsicht in den entsprechenden Sachverhalt (s. Köhler, 1969).

Abzugrenzen ist das sinnvolle Lernen vom mechanischen Lernen (rote learning) (s. Ausubel, Novak, & Hanesian, 1980; Mayer, 2002). Wenngleich das mechanische Lernen eine wichtige Rolle im Unterricht spielt, z.B. beim Lernen von Vokabeln oder historischen Fakten, und oft Grundlage für sinnstiftendes Lernen ist, beschränkt es sich auf die Speicherung von Informationen. Im Gegensatz dazu soll aber beim bedeutungsvollen Lernen, durch die Integration von Informationen, neuer Sinn gestiftet werden (s. Fiorella & Mayer, 2015). Sinnvolles Lernen kann beispielsweise durch konkrete Situationen gefördert werden, in denen den Schülerinnen und Schülern Gelegenheit gegeben wird, ihr Vorwissen zu aktivieren und die neuen Informationen damit abzugleichen. Damit das Gelernte dann wiederum auf weitere Situationen übertragen werden kann, muss es aus dem konkreten Kontext gelöst, also dekontextualisiert werden (Bransford, Franks, Vye, & Sherwood, 1989; Brown, Collins, & Duguid, 1989). Gelingender Transfer stellt daher einen wichtigen Indikator für sinnvolles Lernen dar (s. Mayer, 2002).

**Lernatmosphäre.** Darunter werden jene Aspekte gefasst, die ein förderliches und unterstützendes Lernklima charakterisieren, wie z.B. die Trennung von Lern- und Leistungssituationen, ein konstruktiver Umgang mit Schülerbeiträgen und ein freundliches Auftreten der Lehrperson (s. z.B. Helmke, 2012; Jahn, Prenzel, Stürmer, & Seidel, 2011; Klieme, Lipowsky, Rakoczy, & Ratzka, 2006; Meyer, 2013; Muijs & Reynolds, 2011; Seidel & Shavelson, 2007; Seidel, Prenzel, Rimmele, Schwindt et al., 2006).



**Abb. 3:** Merkmale der Struktur- und Prozessqualität von Unterricht (aus: Plöger & Scholl, 2014c, S. 140)

### Strukturqualität

Die linke Seite der Abbildung stellt Unterricht als *Strukturzusammenhang* dar. Mit Blick auf die beiden großen *didaktischen Theorien*, die bildungstheoretische Didaktik (s. z.B. Klafki, 1995, 2007) und die lerntheoretische Didaktik (s. z.B. Heimann, 1976a, 1976b; Heimann, Otto, & Schulz, 1972), kann die Struktur von Unterricht als Zusammenhang von mindestens vier Elementen gekennzeichnet werden, die das formale Gerüst darstellen, „in dessen Rahmen sich inhaltlich die Planung, Durchführung und *Analyse* von Unterricht bewegt“ (Plöger & Scholl, 2012, S. 241): (1) Ziele/Intentionen/Themen/thematische Struktur, (2) Methoden, (3) Medien sowie (4) Lernvoraussetzungen der Schülerinnen und Schüler. In den entsprechenden Übersichtsdarstellungen bzw. Lehrbüchern der Allgemeinen Didaktik gelten diese Elemente als *unbestritten* (s. z.B. Ditton, 2002; Kron, 2008; Peterßen, 2000; Plöger, 2008; Tulodziecki, Blömeke, & Herzig, 2009).

Gesichert wird die Strukturqualität von Unterricht durch eine *sinnvolle* Abstimmung dieser Elemente. Diese Abstimmung nimmt die Lehrperson bereits bei der Planung von Unterricht vor, indem sie bspw. hinreichend klare und stimmige Lernziele formuliert, aber

auch geeignete Sozialformen, Unterrichtsmethoden und Medien auswählt, die das Erreichen der Ziele unterstützen. In diesem Sinne ist Unterrichtsplanung stets *Strukturplanung*, da es sich um einen „*Entscheidungsprozess* im Strukturzusammenhang von Intention, Inhalt, Methode und Medium (bzw. Medien)“ (Ditton, 2002, S. 200) handelt (s. z.B. Heimann, 1976a; Schulz, 1972, 2006).

### Prozessqualität

Diese Strukturelemente bilden zunächst nur einen *statischen* Rahmen. Im späteren Unterrichtsgeschehen muss die in der Planung entworfene Struktur in einen „*prozessualen* Fluss übersetzt werden“ (Plöger & Scholl, 2014b, S. 276), um auf diese Weise konkrete Lerngelegenheiten zu realisieren, die die Lernprozesse der SchülerInnen optimal unterstützen sollen. Zur Beurteilung dieses Prozessgeschehens sind Kriterien notwendig, die in der Unterrichtsforschung als Merkmale von Unterrichtsqualität bezeichnet werden (s. z.B. Clausen, Reusser, & Klieme, 2003; Clausen, Schnabel, & Schröder, 2002; Ditton, 2002; Einsiedler, 2002; Gruehn, 2000; Helmke, 2012; Kunter et al., 2005; Lipowsky, 2006; Seidel, 2003a; Weinert & Helmke, 1996, 1997; Weinert, Schrader, & Helmke, 1989; Weinert et al., 1990).

Diese Prozessmerkmale von Unterrichtsqualität – gelegentlich werden sie auch als *didaktisch-methodische Prinzipien* oder auch als *Prinzipien effektiven Lehrens* bezeichnet<sup>11</sup> – stellen einen wichtigen Teil pädagogischen Wissens dar, den ich mit Plöger und Scholl (2014c) als *didaktisch-methodisches* Wissen bezeichne.<sup>12</sup> Bei der Analyse von Unterricht fungieren diese Prozessmerkmale bzw. Prinzipien effektiven Lehrens als *Kriterien* zur Beurteilung der Lernwirksamkeit des abgelaufenen Unterrichts (s. z.B. Jamil, Sabol, Hamre, & Pianta, 2015; Krull, Oras, & Pikksaar, 2010; Panasuk & Sullivan, 1998; Wiens, Hessberg, Lo-Casale-Crouch, & DeCoster, 2013).

11 An dieser Stelle sei auf den jeweils unterschiedlichen begrifflich-konzeptuellen Zugriff zwischen „gutem“ und „effektivem“ Unterricht hingewiesen (s. Berliner, 2005; Kunter & Ewald, 2016; Kunter & Trautwein, 2013). Während guter Unterricht (good teaching) eher als Konstruktion vom „grünen Tisch“ und als idealisierte Vorstellungen von Unterricht verstanden werden kann, bezieht sich effektiver Unterricht (effective teaching) auf wirksamen Unterricht, sodass derjenige Unterricht „gut“ ist, für den eine bestimmte (Lern-) Wirksamkeit nachweisbar ist (s. Terhart, 2010, S. 40f.). Berliner (2005) fordert daher eine Vereinigung beider Herangehensweisen zu einem „qualitätvollen Unterricht“. Kunter & Ewald meinen dazu: „Um diesen qualitätvollen Unterricht zu finden, braucht es bestimmte Wertvorstellungen, in die auch theoretische Annahmen über Lernen und die Gestaltung von pädagogischen Settings einfließen. Zur umfassenden Bewertung wird jedoch darüber hinaus die Empirie benötigt, die prüft, ob Unterricht wirklich in der Lage ist, die anvisierten Lernziele zu erreichen“ (Kunter & Ewald, 2016, S. 12).

12 Plöger und Scholl begrenzen das didaktisch-methodische Wissen auf die Prinzipien des effektiven Lehrens, um dadurch „die Handlungen von Lehrpersonen im Unterricht in sehr direkter Weise“ (Plöger & Scholl, 2015, S. 5) beschreiben zu können. Diese Eingrenzung erfüllt zwei Funktionen: Zum einen können die Bestandteile des didaktisch-methodischen Wissens in Handlungskomponenten transformiert werden, die die Lehrperson im Unterricht ein- und umsetzen sollte (s. Plöger & Scholl, 2014b). Unterstellt man nun noch die Erlernbarkeit solcher Handlungskomponenten (s. Baumert & Kunter, 2006), dann sollten sich diese als Handlungskompetenzen einer professionellen Lehrperson ausdrücken, die sie aus ihrem didaktisch-methodischen Wissensfundus heraus generieren kann (s. Plöger & Scholl, 2014b, S. 277). Zum anderen dient das didaktisch-methodische Wissen dazu, das Handeln von Lehrpersonen im Unterricht beschreiben zu können. Im Rahmen der Analysekompetenz müsste eine Lehrperson in der Lage sein, erkennen und beurteilen zu können, ob diese einzelnen Merkmale bzw. Prinzipien effektiven Lehrens umgesetzt wurden und inwieweit diese zu Lernwirkungen im Sinne eines sinnstiftenden Lernens seitens der SchülerInnen beigetragen haben (s. Plöger & Scholl, 2015).

Die *Effektivität* dieser Merkmale ist in einer Vielzahl von Einzelstudien und entsprechenden Metastudien vielfach belegt worden:

- *Zielorientierung* (Bolhuis, 2003; Herweg, 2008; Kyriakides, Christoforou, & Charalambous, 2013; Lipowsky, 2006; Seidel, 2014; Seidel & Shavelson, 2007; Seidel, Blomberg, & Stürmer, 2010),
- *Strukturiertheit* (Brophy, 2000; Good & Brophy, 2000; Helmke, 2007, 2012; Muijs & Reynolds, 2011; Walberg & Paik, 2000; Weinert & Helmke, 1996, 1997),
- *Klarheit* (Helmke, 2007, 2012; Scheerens & Bosker, 1997; Seidel & Shavelson, 2007; Weinert & Helmke, 1996, 1997),
- *Exemplarizität* (Durkin & Rittle-Johnson, 2012; Kalish, Kim, & Young, 2012; Renkl & Atkinson, 2003; Renkl, Atkinson, Maier, & Stanley, 2002; Renkl, Stark, Gruber, & Mandl, 1998),
- *Lebensweltbezug des Themas* (Fey, Gräsel, Puhl, & Parchmann, 2004; Mikelskis-Seifert & Duit, 2010; Muckenfuß, 1995; Parchmann, Ralle, & Demuth, 2000),
- *Verknüpfung mit dem Vorwissen* (Ausubel, Novak, & Hanesian, 1980; Hattie, 2009; Reynolds, 1992; Walberg & Paik, 2000; Wang, Haertel, & Walberg, 1993),
- *Aktivierung zur geistigen Selbsttätigkeit* (Aebli, 1993a, 1993b; Baumert & Kunter, 2013; Kunter & Voss, 2013; Leuders & Holzäpfel, 2011; Lipowsky, 2006; Lipowsky et al., 2009),
- *Aufgabenorientierung* (Jordan et al., 2008; Kleinknecht, Maier, Metz, & Bohl, 2011; Maier, Bohl, Kleinknecht, & Metz, 2013; Neubrand, Jordan, Krauss, Blum, & Löwen, 2011),
- *Diskursivität als Reflexion von Inhalt und Methode* (Cohors-Fresenborg & Kaune, 2003, 2007; Reiss, 2002; Schmidt-Wulffen, 1989),
- *Individualisierung/Differenzierung* (Good & Brophy, 2000; Walberg & Paik, 2000; Waxman, Wang, Anderson, & Walberg, 1985),
- *Vertiefung* (Durcharbeiten) (Aebli, 1993a),
- *Übung und Transfer* (Aebli, 1993a, 1993b; Chi & VanLehn, 2012; Lipowsky, 2006; Renkl, 1998, 2001; Renkl, Hilbert, & Schworm, 2009; Schwartz, Chase, Oppezzo, & Chin, 2011),
- *Ergebnissicherung/Erfolgskontrolle* (Klenck & Schneider, 2006; Voss & Kunter, 2011).

Diese im Modell von Plöger und Scholl angeführten Prozessmerkmale zeigen mehr oder weniger große Schnittmengen mit anderen Forschungsprojekten. Im Rahmen des Projektes *COACTIV* wird z.B. von einer „zieladäquaten Orchestrierung“ (Voss & Kunter, 2011, S. 195) von Unterricht unter besonderer Berücksichtigung der individuellen Lernvoraussetzungen und der Rückmeldung gesprochen (ebd., S. 196ff.). Im Projekt *TEDS-M* werden die Aspekte Strukturierung, Methodenvielfalt, Motivation, Differenzierungsmaßnahmen und Leistungsbeurteilung als zentrale Qualitätsmerkmale von Unterricht betont (s. Blömeke & König, 2010). Und in der *TIMS/III-Studie* (*Third [mittlerweile: Trends] in International Mathematics and Science Study*) wurden diese Merkmale unter „verständnisvollem Lernen“ (Baumert & Köller, 2000, S. 273f.) subsumiert und die Merkmale Vorwissen, Lebensweltbezug, multiple Repräsentationen (*Medieneinsatz*) und die Anwendung des Gelernten als zentrale Aspekte gesehen.

Äquivalente Begriffe zu den von Plöger und Scholl genannten Prozessmerkmalen findet man auch im Projekt *SINUS-Transfer* (*Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts*), welches sich selbst als ein Angebot zur Verbesserung der *Lernwirksamkeit* von naturwissenschaftlichem Unterricht versteht und im Sinne dieser Absicht auch tatsächlich „wirkt“ (s. z.B. Prenzel, Carstensen, Senkbeil, Ostermeier, & Seidel, 2005). Das Projekt besteht aus entsprechenden Modulen wie zielorientiertes und verständnisvolles Experimentieren, Aufgabenkultur, Kompetenzzuwachs sichtbar machen und rückmelden, Sicherung von Basiswissen, kumulatives Lernen, die Lernmotivation stärken usw.

Diese hier aufgelisteten Prozess- bzw. Qualitätsmerkmale von Unterricht werden im Teilkapitel 3.2 – von kleineren Modifikationen abgesehen – letztlich als Kategorien für die Durchführung meiner Qualitativen Inhaltsanalyse genutzt, um das didaktisch-methodische Wissen der ProbandInnen zu erfassen. Dort werden die Kategorien dann auch beschrieben, definiert und mit Ankerbeispielen versehen.

### 1.3 Definition der Facetten des fachlichen (incl. fachdidaktischen) Wissens

Das beschriebene didaktisch-methodische Wissen stellt nur einen Teil der für die Analyse von Unterricht notwendigen Wissensbasis dar. In Ergänzung dazu müssen Lehrpersonen für eine angemessene Analyse von Unterricht auch über entsprechendes fachliches (incl. fachdidaktisches) Wissen (s. Plöger & Scholl, 2014a, 2014b) verfügen. Was darunter zu verstehen ist, erläutere ich in diesem Abschnitt.

In der einschlägigen erziehungswissenschaftlichen wie fachdidaktischen Forschungsliteratur besteht ein breiter Konsens darüber, dass LehrerInnen über ein tiefes Verständnis der von ihnen zu unterrichtenden Fächer verfügen sollten (s. z.B. Ball & Bass, 2000; Baumert & Kunter, 2006, 2011b; Blömeke et al., 2010a, 2010b; Borowski, Neuhaus et al., 2010; Brophy, 1991; Daehler & Shinohara, 2001; Grossmann & Schoenfeld, 2005; Kind, 2009; Lipowsky, 2006; Krauss et al., 2008; Oser & Renolds, 2005; Shulman, 1986, 1987; Terhart, 2002, 2016; van Driel & Berry, 2012). Sobald aber die Frage gestellt wird, welche konkreten Wissensfacetten dazu zu zählen sind, scheint man von einem entsprechenden Konsens mehr oder weniger weit entfernt zu sein. Das gilt nicht nur für das fachliche, sondern auch für das fachdidaktische Wissen, was ich im Folgenden beispielhaft für die Unterrichtsfächer Mathematik und Physik (bzw. für den naturwissenschaftlichen Lernbereich) deutlich machen möchte.

Im Bereich der Mathematikdidaktik werden in den entsprechenden Diskussionen über das *Fachwissen* verschiedene Aspekte betont, wie etwa das Verständnis des Faches, die Berücksichtigung bestimmter Bereiche des Fachwissens oder die Unterscheidung zu beherrschender Anspruchsniveaus der Fachinhalte (s. z.B. Baumert et al., 2010). Hinsichtlich solcher Aspekte scheint ein gewisser Konsens vorzuliegen, wenn man etwa auf die entsprechenden Konzeptualisierungen in Forschungsprogrammen wie *COACTIV* und *TEDS-M* schaut. In beiden Projekten werden fachliche Kompetenzen in den inhaltlichen Bereichen *Arithmetik*, *Algebra*, *Geometrie* und *Stochastik* gemessen. Dabei orientiert man sich

an einem Anspruchsniveau, das von der Sekundarstufe I und II bis hin zur elementaren Universitätsmathematik reicht (s. Krauss et al., 2011, S. 142; Krauss et al., 2008; Blömeke et al., 2009; Döhrmann, Kaiser, & Blömeke, 2010, S. 172).

Im Vergleich dazu scheint es weitaus schwieriger zu sein, ein gewisses Maß an Übereinstimmung hinsichtlich der Modellierung des fachdidaktischen Wissens zu erzielen. Hier herrscht eher eine gewisse Skepsis vor, wenn Krauss et al. bemerken, dass dort „eine Konzeptualisierung des fachdidaktischen Wissens ... nicht auf eindeutige Weise möglich“ (Krauss et al., 2011, S. 137) sei. Diese Situation spiegelt sich in den entsprechenden Konzeptualisierungen in COACTIV und TEDS-M, da in beiden Projekten unterschiedliche Wissensfacetten postuliert werden:

- In COACTIV sind es drei Facetten: *Erklärungswissen* (Verständlichmachen von mathematischen Inhalten), *Wissen über das mathematische Denken von Schülerinnen und Schülern* (mathematikbezogene Schülerkognition, insb. Schülerfehler und Schüler-schwierigkeiten beim Lernen) und *Wissen über mathematische Aufgaben* (kognitives Potential von Mathematikaufgaben) (s. Baumert & Kunter, 2011a, S. 32, 37f.; Krauss et al., 2008, S. 234).
- In TEDS-M werden zwei „Subdimensionen“ als Facetten definiert: *curriculares und planungsbezogenes Wissen* einerseits und *interaktionsbezogenes Wissen* andererseits (s. Döhrmann et al., 2010, S. 175).

Die Diskussion im Bereich der Didaktik der Naturwissenschaften zeigt gewisse Parallelen zu der im Bereich der Mathematikdidaktik. Auch dort wird gefordert, dass Lehrpersonen die weite Spanne zwischen Schulwissen und universitärem Fachwissen beherrschen sollten (s. z.B. Borowski et al., 2010; Borowski, Neuhaus et al., 2010; Riese, 2009; Riese & Reinhold, 2010, 2012; Tepner et al., 2012; van Driel & Verloop, 1999). Die aus dieser Perspektive heraus zu entwickelnden Konsequenzen für die Ausbildung von Lehrpersonen werden aber uneinheitlich diskutiert. So werden z.B. unterschiedliche Positionen hinsichtlich der Frage bezogen, ob dieses Wissen im Studium ausschließlich unter fachsystematischen Gesichtspunkten erworben oder ob es von Anfang an auch schon „pädagogisch überformt“ (Fischler, 2008, S. 30) sein sollte.

Noch deutlicher gehen die Meinungen über das fachdidaktische Wissen auseinander. Insgesamt entsteht der Eindruck, dass „für fachdidaktisches Wissen in den MINT-Fächern bislang weder national noch international eine einheitliche, umfassende und auf ein Rahmenmodell abgestimmte Modellierung“ (Kulgemeyer et al., 2012, S. 2; s. auch Gramzow, Riese, & Reinhold, 2013, S. 8; Riese et al., 2015, S. 57; Borowski & Riese, 2010) vorliegt: „Uneinigkeit besteht nicht nur in der Frage, welche unterrichtsrelevanten Aspekte zum fachdidaktischen Wissen gezählt werden, sondern auch darin, wie sie interpretiert werden“ (Tepner et al., 2012, S. 14; s. dazu auch Gramzow et al., 2013; van Dijk & Kattmann, 2010). Darüber hinaus wird das grundsätzliche Problem der *Eigenständigkeit* des fachdidaktischen Wissens kontrovers diskutiert (s. dazu Abell, 2007, 2008; Park, Chen, & Jung, 2011; van Driel, Verloop, & de Vos, 1998).

Die hier nur in knappen Zügen umrissene Situation hinsichtlich der ungeklärten Frage der fachlichen und fachdidaktischen Kompetenzen (s. ausführlich dazu z.B. Bernholt, 2010; Fischer, Borowski, & Tepner, 2012; Fischer, Borowski, Kauertz, & Neumann, 2010; Park & Chen, 2012; Park & Oliver, 2008) hat in den letzten Jahren zur Entstehung



einer Vielzahl von Arbeiten zur Kompetenzmodellierung und -messung im Bereich der Physikdidaktik geführt (s. z.B. den Überblick in Riese et al., 2015).

Stellvertretend für diese Anstrengungen seien hier zwei Forschungsprogramme exemplarisch angeführt: das Forschungsprogramm *ProwiN* (*Professionswissen von Lehrkräften in den Naturwissenschaften*; Borowski, Neuhaus et al., 2010) und das Programm *ProfiLe-P* (*Professionswissen in der Lehramtsausbildung Physik*, Kulgemeyer et al., 2012; Gramzow, 2015; Riese et al., 2015).

Das übergeordnete Ziel des Forschungsprogramms *ProwiN* (*Professionswissen von Lehrkräften in den Naturwissenschaften*) besteht darin, Wirkungen des Professionswissens von Lehrkräften auf Schülerleistungen zu überprüfen. Dazu möchte die Projektgruppe „das Professionswissen von Lehrkräften in den Fächern Biologie, Chemie und Physik“ (Borowski, Neuhaus et al., 2010, S. 344) untersuchen, damit entsprechende Zusammenhänge zwischen dem Fachwissen, dem fachdidaktischen und dem pädagogischen Wissen beschrieben werden können, um dann darauf aufbauend den „Einfluss dieser drei Dimensionen des Professionswissens auf für Unterrichtsqualität relevante Variablen“ (ebd.), wie z.B. Zielklarheit, kognitive Aktivierung und Lernzuwachs, nachweisen zu können.

Um dem Problem der bisher uneinheitlichen Konzeptualisierungen hinsichtlich der fachlichen Kompetenzen zu begegnen, setzt man in *ProwiN* auf einen *Minimalkonsens*, um „erstmalig aus einer den Naturwissenschaften Biologie, Chemie und Physik gemeinsamen Perspektive“ (Tepner et al., 2012, S.18) ein dreidimensionales Modell des Professionswissens zu entwickeln, wobei der Begriff der *Dimension* ein Äquivalent zum Begriff des *Kompetenzbereiches* im COACTIV-Modell darstellt (s. S. 17):

- Die Dimension *Fachwissen* steht für verschiedene inhaltliche Themenbereiche (im Fach Physik z.B. für die Bereiche Mechanik, Elektrizitätslehre und Optik); *formal* wird zusätzlich zwischen deklarativem, prozeduralem und konditionalem Wissen unterschieden.
- Die Dimension des *fachdidaktischen* Wissens wird im Rückgriff auf Shulman in zwei Facetten unterteilt: Einerseits in das *Wissen über bestimmte fachspezifische Schülerfehler und Schülervorstellungen* (bspw. Präkonzepte oder Misskonzepte [misconcepts], die Schülerinnen und Schüler mit in den Unterricht bringen) und andererseits in *fachspezifische Instruktions- und Vermittlungsstrategien* (die exponierte Stellung sowohl des *Experiments* im Rahmen naturwissenschaftlicher Erkenntnisgenerierung als auch von *Modellen* im naturwissenschaftlichen Unterricht zur Darstellung und Wissensvermittlung).
- Das *pädagogische* Wissen umfasst die Facetten Klassenführung, Unterrichtsmethoden, individuelle Lernprozesse und Leistungsbeurteilung (s. Tepner et al., 2012, S. 17ff.).

Aus diesen drei Dimensionen heraus werden dann Items entwickelt, mit denen das Professionswissen von Lehrpersonen im Bereich Naturwissenschaften erhoben werden soll. Die teilweise noch ausstehenden Ergebnisse dieser Testung werden in Zukunft nähere Aufschlüsse über die Tragfähigkeit der festgesetzten Kompetenzfacetten geben (s. z.B. Kirschner, 2013). In ersten Untersuchungen „konnte man die Unterscheidung der Dimensionen Fachwissen und fachdidaktisches Wissen in einer speziellen Modellierung in der Domäne Physik an einer Stichprobe Studierender empirisch nachweisen“ (Kulgemeyer et al., 2012, S. 5; s. auch Riese & Reinhold, 2010).

In dem *physikspezifischen* Forschungsvorhaben *ProfiLe-P (Professionswissen in der Lehramtsausbildung Physik)* haben sich die beteiligten ForscherInnen das Ziel gesetzt, „ein Rahmenmodell des Professionswissens angehender Physiklehrerinnen und -lehrer sowie von Übungsleiterinnen und -leitern der Physik zu entwickeln“ (Kulgemeyer et al., 2012, S. 1). Ein besonderer Fokus liegt dabei auf der „Modellierung domänenspezifischer und generischer Wissensbereiche, die Lehramtsstudierende der Physik im Hochschulstudium als Teil ihrer professionellen Kompetenz erwerben sollten“ (Riese et al., 2015, S. 56). Für die Realisierung dieser Intention sind drei Teilvorhaben angesetzt, die gleichzeitig eine erste Strukturierungsebene im Rahmenmodell des Professionswissens angehender Physiklehrkräfte darstellen (s. ebd., S. 59):

- 1) DaWis (Fachdidaktisches Wissen von Lehramtsstudierenden und Übungsleitern im Fach Physik): Innerhalb dieses Teilvorhabens „soll ein möglichst umfassendes Modell physikdidaktischer Kompetenz entwickelt werden“ (Kulgemeyer et al., 2012, S. 3), um deklaratives und analytisches fachdidaktisches Wissen modellieren und diagnostizieren zu können.
- 2) EWis (Erklärungswissen als prozedurales fachdidaktisches Wissen): Anknüpfend an das erste Teilvorhaben sollen mit EWis „kompetenzdiagnostische Methoden für die Fähigkeit des Verstehbarmachens physikalischer Sachverhalte bei Physikstudierenden entwickelt werden, um eine diesbezügliche Komponente des professionellen Wissens angehender Physiklehrkräfte zu modellieren“ (ebd.). Eine zentrale Stellung in diesem Vorhaben nimmt die „Standardsituation“ des Erklärens im Unterricht ein, wobei man hinsichtlich dieser Facette prozeduralen fachdidaktischen Wissens davon ausgeht, dass es nicht ausreichend sei, lediglich hohes Fachwissen zu besitzen; man müsse auch in der Lage sein, den entsprechenden Sachverhalt adressatengemäß und sachgerecht erklären zu können.
- 3) FaWis (Fachwissen im Studium zum Lehramt der Physik): Grundlegendes Ziel des dritten Teilvorhabens ist es, „ein Modell für universitäres fachliches Professionswissen und ein Testinstrument zur modellkonformen Erfassung des Fachwissens im Fach Physik“ (ebd., S. 5) zu entwickeln, um dadurch die entsprechend festgelegten Ziele für die Studiengänge Gymnasium/Gesamtschule und Haupt-/Real-/Gesamtschule curricular valide beschreiben zu können.

Diese drei Teilvorhaben bilden gleichzeitig das dem Projekt zugrunde liegende Modell des Professionswissens angehender Lehrpersonen im Fach Physik ab. Die Autorengruppe teilt das Gesamtmodell (Professionswissen) in drei einzelne Modelle auf, die jeweils noch einmal ausdifferenziert werden:

- Das Modell des fachdidaktischen Wissens umfasst vier Facetten: Instruktionsstrategien, Schülervorstellungen, Experimente und Vermittlung eines angemessenen Wissenschaftsverständnisses und fachdidaktische Konzepte.
- Das Modell Erklärungswissen wird in zwei Anforderungsbereiche gegliedert: Adressatengerechtigkeit und Sachgemäßheit einer Erklärung. Es wird angenommen, „dass eine Erklärung ein sprachlicher Akt ist und Erklären den Prozess der adressatengemäßen und sachgerechten Erstellung und Modifikation von Erklärungen bedeutet“ (Riese et al., 2015, S. 61).
- Das Modell Fachwissen wird hinsichtlich dreier Fachstufen unterteilt: Schulwissen, vertieftes Schulwissen und universitäres Wissen.



Die hier skizzenhaft dargestellte Forschungs- und Diskussionslage hat gewisse Schnittmengen in der Diskussion um fachliches und fachdidaktisches Wissen offenbart, aber auch exemplarisch die Schwierigkeiten deutlich gemacht, die im Rahmen der ausgewählten fachdidaktischen Bereiche (Mathematik, Physik) hinsichtlich der Begründung fachlicher und fachdidaktischer Wissensfacetten bestehen.

Für die vorliegende Arbeit sind diese Überlegungen nützlich, um einen gewissen *Minimalkonsens* hinsichtlich fachlichen und fachdidaktischen Wissens in den entsprechenden Projekten offenzulegen. Diesen Minimalkonsens werde ich nutzen, um in Teilkapitel 3.3 die entsprechenden Kategorien für die Qualitative Inhaltsanalyse herzuleiten. Da ich dort dezidiert auf die Herleitung der Wissensfacetten eingehe und sie durch entsprechende Kategorien (= Kompetenzfacetten) operationalisiere, beschränke ich mich an dieser Stelle auf die bloße Nennung dieser Facetten für fachliches (incl. fachdidaktisches) Wissen:

- Physikalische Begriffsbildung
- Durcharbeiten von Fachbegriffen
- Experimentieren als zentrale Fachmethode
- (weitere) fachliche Arbeitsweisen: Phänomene beobachten und genau analysieren; Messdaten grafisch darstellen; kritisches Beurteilen von Messwerten; qualitative Zusammenhänge (ansatzweise) mathematisieren
- Historische Genese physikalischen Wissens
- Produktiver Umgang mit Fehlern
- Lernen als konstruktive Tätigkeit von SchülerInnen

## 1.4 Das Paradigma der Expertiseforschung – Komplexität der Informationsverarbeitung

Die bisherigen Ausführungen haben gezeigt, dass für eine angemessene Analyse von Unterricht didaktisch-methodisches und fachliches (incl. fachdidaktisches) Wissen notwendig ist. Nach Plöger und Scholl (2014a) spielt darüber hinaus aber auch die Fähigkeit eine Rolle, das mehr oder weniger komplexe Zusammenspiel von Situationen und Handlungen erkennen und beurteilen zu können. Zur Begründung dieses Aspekts greifen Plöger und Scholl (2014a) auf die Ergebnisse der Expertiseforschung zurück, die gezeigt hat, dass ExpertInnen in der Lage sind, nicht nur einzelne relevante Situationen und Handlungen zu erkennen, sondern diese auch zu größeren didaktischen Einheiten zu verbinden und darauf aufbauend schließlich den gesamten Unterrichtsprozess in den Blick zu nehmen. Das gelingt ExpertInnen offensichtlich dadurch, dass sie die von ihnen bei der Beobachtung von Unterricht aufgenommenen Informationen in unterschiedlichen Komplexitätsgraden verarbeiten können.

Diese Ergebnisse der Expertiseforschung werde ich im Folgenden in gebotener Kürze referieren, um daran anschließend ein von Plöger und Scholl (2014a) entwickeltes fünfstufiges Modell vorzustellen, mit dem unterschiedliche Grade der Komplexität der Informationsverarbeitung im Rahmen der Analyse von Unterricht klassifiziert werden können. In Anlehnung an dieses Stufenmodell sind dann zur Durchführung meiner Qualitativen Inhaltsanalyse (in Ergänzung zu den erforderlichen Kategorien für didaktisch-methodisches und fachliches Wissen) auch entsprechende Kategorien zur Erfassung unterschiedlicher Grade der Komplexität der Informationsverarbeitung operationalisiert worden.

### 1.4.1 Ergebnisse der Expertiseforschung

In der Expertiseforschung existieren schon seit längerem Beschreibungen des ExpertInnenwissens unter formalen Gesichtspunkten.<sup>13</sup> Die dort stattfindende Diskussion kann hier nur kursorisch wiedergegeben werden und fokussiert vor allem auf jene Aspekte, die für die Frage der Analysekompetenz relevant sind. Die Ergebnisse zeigen, dass sich ExpertInnen von NovizInnen vor allem hinsichtlich ihrer *kognitiven* Tätigkeiten und des *Komplexitätsgrades* zu verarbeitender Informationen unterscheiden (s. z.B. Berliner, 1987, 1992, 2004; Bromme, 2001; Carter, Cushing, Sabers, Stein, & Berliner, 1988; Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987; Gruber, 1994; Krauss, 2011; Palmer, Stough, Burdinski, & Conzales, 2005).

---

13 In der Fachliteratur wird Wissen in formaler Hinsicht in deklaratives und prozedurales Wissen unterschieden. Mit deklarativem Wissen wird meistens konzeptuelles (begriffliches) Wissen bezeichnet. Unter prozeduralem Wissen wird (automatisiertes) Handlungswissen verstanden. Damit werden (teilweise unbewusste) Fertigkeiten wie auch kognitive Fähigkeiten beschrieben, die man benötigt, um entsprechende Handlungen ausführen zu können. Diese Unterscheidung ist für die folgenden Überlegungen keineswegs unwichtig, steht aber eher im Hintergrund, denn mit Gruber muss zunächst einmal konstatiert werden: „Die bloße Unterscheidung zwischen deklarativem und prozeduralem Wissen wird dem Stand der Forschung nicht gerecht, in der längst viel feinere Differenzierungen vorliegen“ (Gruber, 2008, S. 101).

- 1) Die unterschiedlichen kognitiven Tätigkeiten zwischen ExpertInnen und NovizInnen zeigen sich in der Wahrnehmung und Beurteilung von (relevanten) Situationen und Handlungen. NovizInnen sind allenfalls ansatzweise in der Lage, „to selectively attend to the most relevant events ... because they lack sufficient knowledge about the teaching profession and specific subject areas to identify these events“ (Gaudin & Chaliès, 2015, S. 46), weshalb sie meistens einen „inaccurate sense of the effectiveness of the lesson“ (Barnhart & van Es, 2015, S. 85) besitzen.<sup>14</sup> Im Gegensatz dazu können ExpertInnen relevante Situationen identifizieren, die einzelnen Situationen miteinander verknüpfen und so zu größeren Sinneinheiten verdichten. Sie „reconstruct and anticipate the context of the instruction“ (König et al., 2014, S. 78) und können dadurch zwischen zielführenden, d.h. lernwirksamen und weniger lernwirksamen Situationen und Handlungen unterscheiden. Um eine solche Unterscheidung treffen zu können, bedarf es einer *gezielten* Suche nach entsprechenden Informationen und einer *Bewertung* dieser Informationen im Zusammenhang mit der Formulierung von möglichen *Handlungsoptionen*, also alternativen, effektiveren Lernhandlungen (s. Barnhart & van Es, 2015, S. 83; Kersting, 2008, S. 847; König et al., 2014, S. 78; Plöger et al., 2015, S. 167; Sabers, Cushing, & Berliner, 1991).<sup>15</sup>
- 2) Des Weiteren gibt es zwischen ExpertInnen und NovizInnen Unterschiede hinsichtlich des Komplexitätsgrades, mit dem Informationen verarbeitet werden. Bei der Analyse von Unterricht gehen NovizInnen „step by step“ (König et al., 2014, S. 78), d.h. chronologisch (sequentiell) und additiv vor, sodass eingehende Informationen nacheinander abgearbeitet werden. Dabei neigen sie dazu, „to focus on superficial matters such as teacher and student characteristics, fleeting classroom management issues and global judgements of lesson effectiveness“ (Castro, Clarke, Jacobs, & Givvin, 2005, S. 11). Dadurch, dass sich die Wahrnehmung von NovizInnen meistens auf einzelne Situationen beschränkt, sind sie auch kaum in der Lage, den Stellenwert von einzelnen Situationen und/oder Handlungen vor dem Hintergrund des gesamten Unterrichtsverlaufes zu beurteilen. Diese Leistung erbringen aber ExpertInnen: Sie können den tieferliegenden didaktisch-methodischen Sinn (Tiefenstrukturen) des unterrichtlichen Handelns interpretativ erschließen. Das gelingt ihnen, weil sie, im Vergleich zu NovizInnen, eine Vielzahl kohärenter Beziehungsstiftungen in Form von Schemata besitzen, die zu semantischen Netzwerken bzw. mentalen Modellen verknüpft sind (Baumert & Kunter, 2006, S. 483; de Jong & Ferguson-Hessler, 1996, S. 111; Gruber & Gallenberger, 2006, S. 26; Shuell, 1990, S. 542; Strahan, 1989, S. 54; Elio & Scharf, 1990). Dadurch sind ExpertInnen in der Lage, eingehende Informationen differenzierter mit dem verfügbaren Wissensnetz zu enkodieren (Berliner 1992, S. 243; Livingston

14 Im deutschsprachigen Raum wird für die Wahrnehmung sichtbarer Situationen und Handlungen der Begriff der *Sichtstrukturen* verwendet (Aebli, 1971; Klieme, 2006; Kunter & Trautwein, 2013; Oser & Baeriswyl, 2002; Oser & Patry, 1994; Reyer, 2004; Seidel, 2003a, 2006b).

15 Bei dieser Suche zeichnen sich ExpertInnen durch ihre hohe Problemlösekompetenz, ihre Entscheidungsstärke und ihr Urteilsvermögen aus (s. z.B. Gruber, Harteis, & Rehrl, 2006, S. 195; Hattie, 2003, S. 6; Shavelson, 1973; Shavelson & Stern, 1981). Damit diese Fähigkeiten hinreichend ausgeprägt sind, bedarf es umfangreicher Erfahrungen und Übung, denn nur so können auf lange Zeit gesehen entsprechende *prozedurale Routinen* ausgebildet werden (Bromme, 1985, S. 184; Bromme & Haag, 2008, S. 807). Diese Routinen sind kognitiv und emotional für das Arbeitsgedächtnis entlastend (Chi, 2006, S. 24; Leinhardt & Greeno, 1986, S. 95).

& Borko, 1990) und tiefer zu verarbeiten, wie dies auch bei der Analyse von Unterricht der Fall ist.<sup>16</sup>

Diese beiden zentralen Erkenntnisse der Expertiseforschung im Hinblick auf die Analyse von Unterricht werden auch in gegenwärtigen Modellierungen zur Analysekompetenz berücksichtigt, wie ich anhand von drei Untersuchungen zeigen möchte.

In den Untersuchungen von van Es und Sherin wird zur Unterscheidung der Analysefähigkeit von Personen ein Dreischritt, bestehend aus den kognitiven Tätigkeiten *noticing*, *interpreting* und *informing decision*, beschrieben (van Es & Sherin, 2002, 2008).<sup>17</sup> *Noticing* meint, Situationen und Handlungen wahrzunehmen und zu beschreiben. Unter *interpreting* verstehen die Autorinnen, dass Personen in der Lage sind, einzelne Situationen miteinander in Beziehung zu setzen und deren Lernwirksamkeit zu bewerten. *Informing decision* heißt, dass vor dem Hintergrund der festgestellten Lernwirksamkeit alternative Handlungsoptionen (pedagogical solutions) entwickelt werden (s. van Es & Sherin, 2002, S. 582).

Die Forschungsgruppe um Seidel (s. z.B. Seidel & Stürmer, 2014) greift auf die Dreiteilung von van Es & Sherin zurück und entwickelte das Konzept der „professionellen Unterrichtswahrnehmung“.<sup>18</sup> Dabei unterscheiden sie drei kognitive Tätigkeiten voneinander, mit denen unterschiedliche Dimensionen professioneller Unterrichtswahrnehmung bezeichnet werden: Beschreibung, Erklärung und Vorhersage (s. z.B. Seidel & Stürmer, 2014). Auf der untersten Stufe (Beschreiben) müssen „lehr-lernrelevante beobachtete Unterrichtssituationen“ (Jahn, Stürmer, Seidel, & Prenzel, 2014, S. 172) *beschrieben* werden. Auf der nächsthöheren Stufe (Erklären) müssen diese Beschreibungen „auf Basis wissenschaftlicher Theorien und Befunde vor dem Hintergrund von Lernwirksamkeit“ (ebd.) *erklärt* werden. Diese Erklärungen sollten dann auf der höchsten Stufe (Vorhersage) genutzt

16 Dabei kann sich der Informationsgehalt von semantischen Netzen hinsichtlich Umfang und Abstraktionsgrad deutlich unterscheiden, denn semantische Netze sind ein komplexes Gebilde von Teil-Ganzes-Strukturen und Mikro- und Makropropositionen, die aus einzelnen Schemata bestehen, die selbst wieder in umfassende Netze integriert sind. In der Literatur wird dann von *tief verstandenem* Wissen gesprochen, wenn die unterschiedlich komplexen Strukturen einer Person deutliche Verbindungen aufweisen und diese idealiter in Form von übergeordneten *Prinzipien* (Peterson & Comeaux, 1987, S. 327) oder in Form von *Abstraktionshierarchien* gespeichert sind (Berliner, 1987, S. 298; Frey, 2004; Leinhardt & Smith, 1985). Durch solche Abstraktionshierarchien besitzen ExpertInnen eine *Beweglichkeit* im Denken, die sie befähigt, über den Prozess des *Chunking*, mit einer vergleichsweise kleinen Anzahl hoch abstrakter Schemata Realität mehrperspektivisch wahrzunehmen (Borko & Livingston, 1989, S. 490). Diese Beweglichkeit offenbart sich aber nicht nur in inhaltlicher Hinsicht (Perspektivenwechsel), sondern äußert sich auch in der Fähigkeit, zwischen verschiedenen Ebenen (abstrakt vs. konkret, strukturell vs. funktional oder oberflächen- vs. prinzipienorientiert) je nach situativer Anforderung strategisch wechseln zu können (s. Krems, 1996, S. 83).

17 Dieser Einteilung von van Es und Sherin wird auch in anderen Studien gefolgt. Beispielsweise greifen König et al. (2014) auf diese Einteilung zurück, differenzieren aber nur zwischen zwei unterschiedlichen kognitiven Tätigkeiten, mit denen Informationen bei der Analyse von Unterricht verarbeitet werden: *noticing* und *interpreting*. Sie versuchen dann, Korrelationen zwischen dem pädagogischen Wissen und den kognitiven Tätigkeiten nachzuweisen.

18 Mit dem Konzept der „professionellen Unterrichtswahrnehmung“ wird beschrieben, „wie Lehrpersonen komplexe Unterrichtssituationen vor dem Hintergrund ihres Wissens um Lehren und Lernen beobachten und interpretieren“ (Jahn et al., 2014, S. 172). Zur Messung der professionellen Unterrichtswahrnehmung wird das Tool Observer (Seidel & Stürmer, 2014; Seidel, Blomberg, & Stürmer, 2010) eingesetzt, das ein inhaltlich valides, videobasiertes Instrument zur Erfassung der professionellen Wahrnehmung von Unterricht darstellt (s. z.B. Blomberg, Stürmer, & Seidel, 2011; Jahn et al., 2014; Seidel et al., 2010; Seidel et al., 2011; Stürmer & Seidel, 2015).

werden, um „Auswirkung auf weitere Lernprozesse von Schülerinnen und Schülern“ (ebd.) *vorhersagen* zu können.

Kersting (2008) und Kersting et al. (2012) unterscheiden drei Grade (Level) der Komplexität der Informationsverarbeitung, in denen die Tiefe der Interpretation zum Ausdruck kommt: Level 1 bezieht sich auf die (reine) Beschreibung von beobachteten Unterrichtsergebnissen. Level 2 umfasst Aussagen, die zwar analytische Schlussfolgerungen enthalten, aber nicht miteinander in Beziehung gesetzt werden, sodass sie (noch) kein zusammenhängendes Urteil bilden können („but without connecting analytical points to form a coherent argument“, Kersting, 2008, S. 849). Personen, die Level 3 erreicht haben, geben Antworten, die eine sehr umfassende, kohärente und integrierte Interpretation (hoher Komplexitätsgrad der Informationsverarbeitung) darstellen und sich durch das Erkennen von Ursache-Wirkungsbeziehungen (kognitive Tätigkeit) auszeichnen (s. Kersting, 2008; Kersting et al., 2012).

### 1.4.2 Fünfstufiges Kompetenzmodell nach Plöger und Scholl

Die herangezogenen Studien belegen, dass bei der Analyse von Unterricht Informationen in unterschiedlichem Komplexitätsgrad zu verarbeiten sind und dabei noticing, interpreting und decision making wichtige Facetten von Analysekompetenz darstellen.

Diese beiden Aspekte (die bei der Analyse von Unterricht zu unterscheidenden kognitiven Tätigkeiten des Wahrnehmens, Evaluierens und des Suchens nach Alternativen einerseits und die Komplexität der Informationsverarbeitung andererseits) berücksichtigen auch Plöger und Scholl in ihrem Modell der Komplexität der Informationsverarbeitung. Sie sehen die Analyse von Unterricht als einen Prozess der Informationsverarbeitung und konkretisieren diesen Prozess durch fünf Stufen.<sup>19</sup>

<sup>19</sup> Bei der Modellierung ihres Modells greifen Plöger und Scholl (2014a) auf das *Modell of hierarchical complexity* von Commons et al. zurück. Dieses Modell von Commons et al. (s. z.B. Commons et al., 1982; Commons & Richards, 1984a, 1984b; Commons & White, 2003; Commons, Goodheart, Dawson, & Draney, 2008; Commons, Trudeau, Stein, Richards, & Krause, 1998) ist *inhaltsunabhängig*, also *formal* gehalten, und erhebt somit den Anspruch, auf sämtliche Inhaltsbereiche angewendet werden zu können, in denen *Informationen in organisierter Form verarbeitet* werden, wie dies auch auf die Analyse von Unterricht zutrifft. Organisierte Informationsverarbeitung ist immer dann erforderlich, „wenn mit zunehmender Anzahl von Informationselementen die Schwierigkeit ihrer Verarbeitung steigt“ (Plöger et al., 2015, S. 169). Dies ist dann der Fall, wenn die Anzahl der zu verbindenden Elemente größer wird und dadurch der Komplexitätsgrad der zu lösenden Aufgabe steigt. Um die *Art* der dazu erforderlichen Komplexität der Informationsverarbeitung besser beschreiben zu können, unterscheiden Commons et al. (s. z.B. Commons et al., 2008) zwischen *horizontaler* und *vertikaler* Komplexität.

Die *horizontale* Komplexität ist dadurch bestimmt, dass es lediglich zu einer willkürlichen Verkettung (chains) von Elementen kommt, sodass die zu verarbeitenden Informationen mehr oder weniger unverbunden bleiben. Im Gegensatz dazu ist die *vertikale* Komplexität nicht nur durch die steigende Anzahl von Elementen gekennzeichnet, sondern dadurch, dass zwischen diesen Elementen spezifische Verbindungen hergestellt werden. Diese Verbindungen sind keine nichtwillkürlichen (nonarbitrary) Verkettungen, sondern stellen eine *Koordination* dar (Commons, 2007), sodass weniger komplexe Einheiten in höheren organisiert werden und dadurch unterschiedliche Komplexitätsgrade der Informationsverarbeitung durch eine geordnete Abfolge von Stufen modelliert werden können. Die einzelnen Stufen bauen aufeinander auf, sodass sich höhere Stufen aus der Koordination von Elementen der darunterliegenden ergeben und über die Stufen hinweg die Komplexität in *quantitativer* wie *qualitativer* Hinsicht zunimmt. Das bedeutet, dass komplexere Aufgaben (= höhere Stufen des Modells) Leistungen voraussetzen, die für eine Stufe geringerer Komplexität kennzeichnend sind, sodass sich komplexere Aufgaben nur durch *Rekursion* (Commons et al., 2008, S. 183) auf weniger komplexe Subaufgaben (= niedrigere Stufen des Modells) lösen lassen.

Das Modell ist in *drei* Hauptstufen (römische Bezifferung) und *fünf* Unterstufen (arabische Bezifferung) unterteilt:

### Hauptstufe I: Analytische Kompetenz

Auf Stufe 1 und 2 sind die Analyseleistungen begrenzt auf die Erfassung *einzelner* Unterrichtsereignisse, die als sichtbare Handlungen und/oder Situationen wahrgenommen werden, sodass mit diesen beiden Stufen die *Sichtstruktur* (Aebli, 1971; Klieme, 2006; Kunter & Trautwein, 2013; Oser & Baeriswyl, 2002; Oser & Patry, 1994; Reyer, 2004; Seidel, 2003a, 2006b) von Unterricht abgebildet wird. Die Bezeichnung „Analytische Kompetenz“ betont dabei, dass die ProbandInnen in der Lage sind, *einzelne* Ereignisse aus dem Unterrichtsgeschehen herausfiltern und beschreiben zu können. Solche Leistungen können auch schon NovizInnen erbringen, wie die Auswertung des Tests (s. Kap. 1.5.2, S. 40) gezeigt hat. NovizInnen waren in der Lage, Situationen und Handlungen wahrzunehmen (Stufe 1) und auch den damit verbundenen Grad der Ausprägung (z.B. hohe vs. niedrige Motivation der SchülerInnen; hoher vs. niedriger Gesprächsanteil der Lehrperson) zu registrieren (Stufe 2).

Stufe III: Prozesskompetenz		beziehen sich auf die <b>Tiefenstrukturen</b> von Unterricht
	Experte	
Stufe 5	– kann den Stellenwert von didaktisch-methodischen Sinneinheiten vor dem Hintergrund des <i>gesamten</i> Unterrichtsprozesses einschätzen und für solche Einheiten begründbare Alternativen angeben	
Stufe II: Synthetische Kompetenz		beziehen sich auf die <b>Tiefenstrukturen</b> von Unterricht
Stufe 4	– kann <i>einzelne</i> Handlungen oder Situationen zu <i>größeren</i> didaktisch-methodischen Einheiten verdichten und für solche Einheiten Alternativen angeben	
Stufe 3	– kann die Effekte bzw. Folgen, die <i>einzelne</i> Handlungen oder Situationen bewirken (können), identifizieren oder mit hoher Wahrscheinlichkeit kalkulieren bzw. voraussehen	
Stufe I: Analytische Kompetenz		beziehen sich auf die <b>Sichtstrukturen</b> von Unterricht
Stufe 2	– kann den Grad der Ausprägung dieser Kategorien bestimmen	
Stufe 1	– kann einzelne Unterrichtsereignisse (Handlungen oder Situationen) in entsprechende Kategorien einordnen	

Abb. 4: fünfstufiges Modell der Analysekompetenz (aus: Plöger et al., 2015, S. 170)

## Hauptstufe II: Synthesekompetenz

Ab Stufe 3 aufwärts werden Analyseleistungen beschrieben, die die *Tiefenstrukturen* von Unterricht betreffen. Damit solche *Tiefenstrukturen* (Kunter & Trautwein, 2013; Oser & Patry, 1994; Reyer, 2004; Seidel, 2003a, 2003b) erfasst bzw. erschlossen werden können, müssen die ProbandInnen die sichtbaren Handlungen und/oder Situationen in einem *interpretativen* Akt zu didaktisch-methodischen Sinnzusammenhängen verdichten können (s. z.B. König & Lebens, 2012). Mit den Stufen 3 und 4 werden Grade der Analysekompetenz beschrieben, bei denen einzelne Handlungen mit weiteren einzelnen Handlungen oder mit dadurch hervorgerufenen Effekten zu größeren didaktisch-methodischen Einheiten *verbunden* werden und zu diesen Alternativen formuliert werden können. Bei ProbandInnen, die Leistungen erbringen, die die Stufen 3 und 4 betreffen, ist die Informationsverarbeitung deutlich komplexer organisiert als auf den beiden unteren Stufen. Durch das Zusammenfassen (synthetisieren) einzelner Aspekte werden diese beiden Stufen von Plöger und Scholl als „Synthesekompetenz“ bezeichnet.

## Hauptstufe III: Prozesskompetenz

Eine Person, die Stufe 5 erreicht hat, bezieht sich auf den *gesamten* zu analysierenden Unterrichtsprozess. Sie kann didaktisch-methodische Sinneinheiten vor dem Hintergrund des *gesamten* Unterrichtsverlaufs beurteilen und sowohl für solche Einheiten als auch für den gesamten Unterrichtsprozess begründbare Alternativen angeben. Daher wird diese Stufe auch als „Prozesskompetenz“ benannt, da hier der komplette Unterrichtsverlauf erfasst werden muss.

Das hier in aller Kürze vorgestellte fünfstufige Kompetenzmodell von Plöger und Scholl wird auch bei meiner Qualitativen Inhaltsanalyse zur Erfassung der Komplexität der Informationsverarbeitung genutzt. Die zum Einsatz kommenden Kategorien operationalisiere ich in Kapitel 3 und nenne an dieser Stelle die Kategorien, die zur Auswertung des Interviewmaterials zum Einsatz kommen werden:

- Verstehenstiefe – Level 1
- Verstehenstiefe – Level 2a
- Verstehenstiefe – Level 2b
- Verstehenstiefe – Level 3a
- Verstehenstiefe – Level 3b



## 1.5 Methode und Ergebnisse der Untersuchung von Plöger und Scholl (2014)

### 1.5.1 Anlage der Studie

Das im vorangegangenen Teilkapitel vorgestellte fünfstufige Kompetenzmodell wurde von Plöger und Scholl zur Konstruktion eines Tests genutzt, mit dem die Analysekompetenz von Lehrpersonen gemessen wurde.<sup>20</sup> Dieser Test (paper and pencil-Verfahren) bestand aus 23 Items (Plöger & Scholl, 2014b, S. 273), zu deren Beantwortung die ProbandInnen didaktisch-methodisches und fachliches (incl. fachdidaktisches) Wissen aktivieren mussten, um so die *Lernwirksamkeit* des unterrichtlichen Handelns angemessen einschätzen zu können. Exemplarisch soll hier ein Item vorgestellt werden, für dessen Lösung didaktisch-methodisches Wissen (hier Exemplarizität [Einheit von Konkretem und Abstraktem]) notwendig ist (s. Plöger & Scholl, 2014b, S. 280): „Zu Beginn der Unterrichtsstunde lässt der Lehrer Beispiele für Lichtbrechung nennen und führt dann selbst weitere an (mittels bildlicher Darstellung). Wie beurteilen Sie diesen Einstieg?“

Den ProbandInnen wurde als Stimulus eine videographierte Unterrichtsstunde vorgeführt (11. Klasse, Fachoberschule, Fach Physik; Thema: Snelliussches Brechungsgesetz). In dieser Physikstunde sollten die SchülerInnen das Brechungsgesetz nachentdecken: Zu Beginn der Unterrichtsstunde präsentiert die Lehrperson verschiedene Phänomene, in denen Brechung auftritt, in Form von Bildern. Im Anschluss daran werden die SchülerInnen gebeten, ein Experiment zu planen, mit dem das Brechungsgesetz nachentdeckt werden kann. Die SchülerInnen führen dann das Experiment durch, werten ihre Ergebnisse aus und formulieren mit Hilfe des Lehrers das Snelliussche Brechungsgesetz, das mittels Beispielrechnung (Berechnung des Brechungsindex aus gegebenen Winkeln) überprüft wird. Zum Abschluss der Unterrichtsstunde demonstriert die Lehrperson den SchülerInnen die Funktionsweise einer Fresnel-Linse und deklariert diese Linse als „Anwendung“ des Brechungsgesetzes.

Dieses Video wurde vornehmlich aus zwei wichtigen Gründen eingesetzt:

- 1) Das Video ist insofern repräsentativ, weil es typische Schwächen von Physikunterricht im Sinne der IPN-Videostudie (*Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik*, Seidel & Prenzel, 2007; Seidel, Prenzel, Rimmele, Schwindt et al., 2006; Seidel, Prenzel, Rimmele, Dalehefte et al., 2006) zeigt. Negativ fallen vor allem „die fehlende Einbettung des Experiments in einen hypothesengeleiteten Prozess, die starke Lehrerzentrierung, die daran gekoppelte reproduktive Gesprächsführung, die mangelhafte Phasierung der Stunde unter dem Gesichtspunkt der Lernaktivitäten und die geringe explizite und implizite Zielorientierung auf“ (Plöger & Scholl, 2014a, S. 93).
- 2) Da das Video eine komplette Unterrichtsstunde zeigt, ist die Komplexität der wahrnehmbaren bzw. erschließbaren Situationen und Handlungen entsprechend groß, so dass genügend Äußerungs- und Anknüpfungsmöglichkeiten für die ProbandInnen entstehen. Dieses Maß an Komplexität stellt zusammen mit den Schwächen des gezeigten

<sup>20</sup> Die Studie war als Querschnittsstudie angelegt. Die Daten wurden im Zeitraum von Juli 2010 bis März 2011 erhoben.



Unterrichts insbesondere für ExpertInnen einen hohen Anreiz dar, mögliche *Handlungsalternativen* zu entwickeln (s. Gruber, 1994).

Die Gesamtstichprobe (N = 800) umfasste vier Teilgruppen mit je 200 Personen, die an die berufsbiographischen Entwicklungsmöglichkeiten im Bereich der LehrerInnenbildung angelehnt waren: Lehramtsstudierende (an Universitäten aus sechs Bundesländern), ReferendarInnen (an Studienseminaren aus vier Bundesländern), LehrerInnen (an Gymnasien eines regionalen Bezirks) und Fach- und SeminarleiterInnen (an Studienseminaren aus sechs Bundesländern).<sup>21</sup>

Um überprüfen zu können, inwieweit das *fachliche* (incl. *fachdidaktische*) Wissen Einfluss auf die Analyseleistungen von Unterricht hat, wurde die Stichprobe zusätzlich hinsichtlich der Fachzugehörigkeit der ProbandInnen in drei Gruppen ausdifferenziert:

- Personen, die das Fach Physik unterrichten bzw. studieren,
- Personen, die mindestens ein naturwissenschaftliches Fach und/oder das Fach Mathematik unterrichten bzw. studieren,
- Personen, die weder ein naturwissenschaftliches Fach und auch nicht das Fach Mathematik unterrichten bzw. studieren.

**Tab. 1:** Zusammensetzung der Stichprobe nach Fachzugehörigkeit (in Anlehnung an Plöger & Scholl, 2014b, S. 280)<sup>22</sup>

	Studierende	Referendare	Lehrpersonen	Fach- & SeminarleiterInnen
Physik als Unterrichtsfach	8	7	36	20
Mathematik und/oder mindestens ein naturwissenschaftliches Fach	90	40	56	58
Kein naturwissenschaftliches Fach	102	153	108	122

Der Untersuchung von Plöger und Scholl lagen zwei *Hypothesen* zugrunde (s. Plöger & Scholl, 2014a, 2014b):

- Hypothese 1: „*Experten (Fach- und Seminarleiter) erreichen die höchsten Stufen der Analysekompetenz, während die Kompetenz von Novizen (Studierenden) über die ersten zwei Stufen nicht weit hinausgeht.*“<sup>23</sup> (Plöger & Scholl, 2014b, S. 279, kursiv im Original)
- Hypothese 2: „*Personen, die das Fach Physik oder Mathematik bzw. ein weiteres naturwissenschaftliches Fach unterrichten, erzielen aufgrund ihrer fachlichen bzw. fachdidaktischen Kompetenzen höhere Analyseleistungen als Personen, die keines dieser Fächer unterrichten.*“ (Plöger & Scholl, 2014b, S. 279, kursiv im Original)

21 Die Stichprobe für die Studierenden, ReferendarInnen und Fach- und SeminarleiterInnen wurde als Gelegenheitsstichprobe gezogen, die als Klumpen erhoben wurde. Die Gruppe der Lehrpersonen wurde als Klumpenstichprobe durch eine zweistufige Zufallsauswahl erhoben.

22 Zur Harmonisierung hinsichtlich der Stichprobenzusammensetzung wurde die ursprüngliche Tabelle von Plöger und Scholl (2014b) angepasst.

23 Als Rangfolge hinsichtlich der Stichprobe wird dabei erwartet, dass Fach- und SeminarleiterInnen die höchsten Stufen der Analysekompetenz erreichen. Ihnen sollten LehrerInnen und ReferendarInnen folgen. Die niedrigste Ausprägung der Analysekompetenz sollten Studierende aufweisen.

### 1.5.2 Zentrale Ergebnisse der Studie von Plöger und Scholl (2014)

Die 23 Items bildeten eine raschhomogene Skala mit einer Reliabilität von .79 (Plöger & Scholl, 2014a).<sup>24</sup> Die nachfolgenden Ergebnisse stellen die Ergebnisse im Hinblick auf die angeführten Hypothesen dar (s. Plöger & Scholl, 2014a, 2014b, 2014c, 2016):

- Hinsichtlich der Höhe der Analysekompetenz gibt es deutliche Unterschiede zwischen den ProbandInnengruppen: Studierende erzielen einen relativ niedrigen Fähigkeitswert, Fach- und SeminarleiterInnen dagegen den höchsten. Zwischen diesen beiden Gruppen liegen die ReferendarInnen und die LehrerInnen.<sup>25</sup>
- Die Studierenden können (von wenigen Ausnahmen abgesehen) nur die Items der ersten beiden Unterstufen des fünfstufigen Modells lösen. Die Fach- und SeminarleiterInnen erreichen durchschnittlich die höchsten Leistungen und sind in der Lage, auch Items zu lösen, die der fünften Stufe des Modells zuzuordnen sind.
- Darüber hinaus konnte nachgewiesen werden, dass das fachliche Wissen eine zentrale Voraussetzung für die Analyse der videografierten Unterrichtsstunde ist, denn Personen, die das Fach Physik unterrichteten, erreichten signifikant höhere Analyseleistungen als Personen, die dieses Fach nicht unterrichten ( $[F(2, 788) = 5,973, p < 0,01]$ ). Allerdings gab es unter den besten 30 Personen auch „Ausreißer“, also „auch einige wenige, die exzellente Analyseleistungen erzielen, obwohl sie weder Physik noch Mathematik und/oder ein anderes naturwissenschaftliches Fach unterrichten.“ (Plöger & Scholl, 2016, S. 183)

---

24 Da sich die Itemskala als *raschhomogen* (Reliabilität von .79, s. Plöger & Scholl, 2014a) erwies, können die vorliegenden Summenwerte im Raschmodell als eine *suffiziente* Statistik gewertet werden. Der Grundgedanke dabei ist, dass die gelösten Aufgaben als Indikator für die Personenfähigkeit herangezogen werden: „Eine Person mit einer niedrigen Ausprägung dieser Fähigkeit wird z.B. nur die einfachen Aufgaben eines Tests lösen können. Eine Person mit einer höheren Fähigkeit hingegen wird sowohl die einfachen als auch die schwierigen Aufgaben lösen können – und damit insgesamt mehr Aufgaben lösen.“ (Strobl, 2010, S. 15)

25 Die über die Varianzanalyse ermittelten Leistungen der vier ProbandInnengruppen unterschieden sich signifikant voneinander ( $F[3, 796] = 84,554, p < .001$ ). Eine anschließende Post-Hoc-Analyse (Tukey-HSD-Test) zeigte, dass sich Studierende signifikant von allen anderen Gruppen unterschieden ( $p < .01$ ). „Dagegen unterscheiden sich die Lehrpersonen nur von den Studierenden signifikant. Referendarinnen und Referendare und Seminar-/Fachleiterinnen und -leiter unterscheiden sich signifikant sowohl von den Studierenden als auch voneinander, allerdings nicht von den Lehrpersonen.“ (Plöger & Scholl, 2014a, S. 104)

## 1.6 Quantitativer Validierungsversuch von Plöger, Scholl und Seifert (2015)

Im Sinne der oben referierten Erkenntnisse aus der Expertiseforschung waren diese Ergebnisse zu erwarten (s. Kap. 1.4.1). Insbesondere der große Unterschied zwischen den Leistungen der Fach- und SeminarleiterInnen (ExpertInnen) und der Studierenden (NovizInnen) war ein Beleg dafür, dass der Test eine klare Unterscheidung von „Extremgruppen“ zulässt, von denen vorab angenommen werden konnte, dass die eine Gruppe das in Frage stehende Merkmal (hier Analyseleistung) in hohem Maße besitzt, während dies bei der anderen Gruppe gerade nicht der Fall ist.

Im Sinne eines solchen *well-known-group-Vergleichs* (Cronbach & Meehl, 1955; DeVellis, 1991; Hattie & Cooksey, 1984; Rubin & Babbie, 2015) stellte dieses Testergebnis somit einen wichtigen Beleg für die Validität des zugrunde liegenden Konstrukts Analysekompetenz dar.<sup>26</sup> Um dieser Frage der Validität des Konstrukts weiter nachzugehen, führten Plöger et al. (2015) zusätzlich Konfirmatorische Faktorenanalysen (CFA) durch.

Im Folgenden erläutere ich diesen quantitativen Validierungsversuch eingehend, weil er den *zentralen* Bezug zu meiner Studie darstellt, in der ich auf qualitativem Wege prüfen möchte, ob die von mir erhobenen empirischen Daten ebenfalls mit den Annahmen des Konstrukts interpretiert werden können.

### 1.6.1 Modellprüfung mittels Konfirmatorischer Faktorenanalysen

Zunächst greife ich auf die Grafik aus der Einleitung (s. Abb. 1, S. 15) zurück und nenne noch einmal die zentralen Annahmen, die das Konstrukt Analysekompetenz bestimmen: Das Konstrukt Analysekompetenz wird über zwei Dimensionen, eine inhaltliche und eine formale, mit entsprechenden Faktoren definiert. Die inhaltliche Dimension ist dabei durch zwei Faktoren näher bestimmt: das *didaktisch-methodische* (DM) und das *fachliche* (incl. *fachdidaktische*) (F) Wissen. Die formale Dimension wird über den Grad der für die Analyseleistungen notwendigen Komplexität der Informationsverarbeitung abgebildet. In Anlehnung an das fünfstufige Kompetenzmodell (s. Abb. 4, S. 36) ist die formale Dimension dabei durch drei Faktoren gekennzeichnet (in aufsteigender Reihenfolge), die den oben beschriebenen Hauptstufen des Modells entsprechen: Die Stufen 1 und 2 des Modells repräsentieren den Faktor der *Analytischen Kompetenz* (AK), die Stufen 3 und 4 stellen die *Synthetische Kompetenz* (SK) dar und die Stufe 5 steht für den Faktor der *Prozesskompetenz* (PK).

Je nach Wahl bzw. Kombination dieser Dimensionen und der zugehörigen Faktoren ergeben sich mindestens fünf mögliche Strukturmodelle, die das Konstrukt repräsentieren und deren Validität im Rahmen der CFA geprüft wurde (s. Plöger et al., 2015):

---

<sup>26</sup> Eine zentrale Rolle hatte die Frage der *inhaltlichen* Validität bereits bei der Testkonstruktion und -auswertung gespielt. Die inhaltliche Validität der Testitems wurde über ein Rating von ca. 40 Seminar- und FachleiterInnen gesichert. Diese Personen waren an der Diskussion und Prüfung der akzeptierbaren Antworten auf die einzelnen Items beteiligt. Nach Abschluss der Diskussion konnte ein *Auswertungsmanual* erstellt werden, das die Grundlage für die Kodierung der Antworten der ProbandInnen bildete.

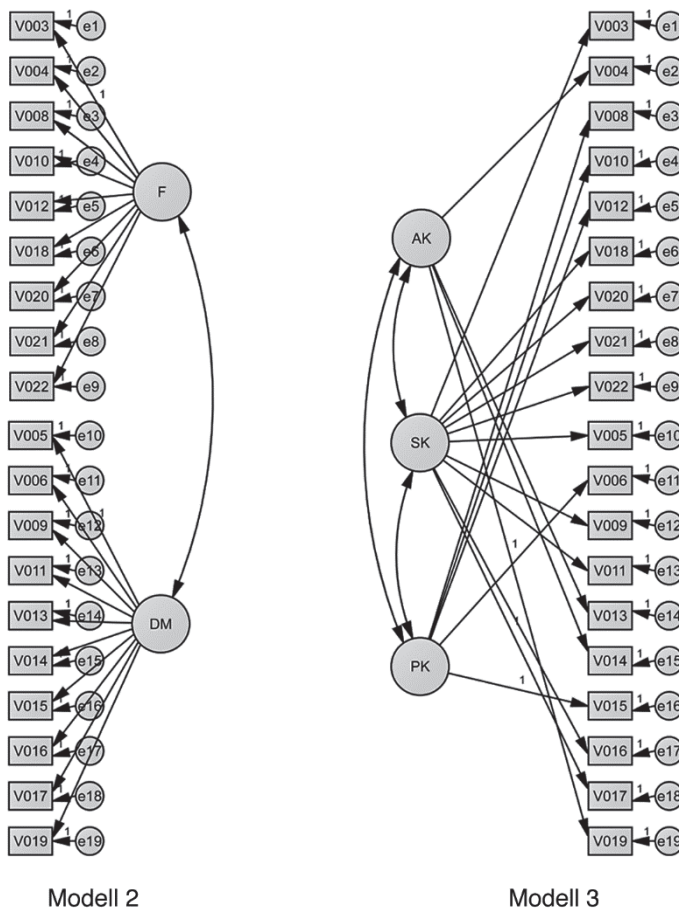
*Modell 1: eindimensionales Konstrukt (Analysekompetenz als g-Faktor)*

Mit diesem Modell wird die Varianz aller manifesten Variablen durch einen Generalfaktor erklärt. Aufgrund der einfachen Struktur des Modells hatten Plöger et al. (2015) auf eine Abbildung verzichtet.

*Modell 2* (s. Abb. 5, S. 42): *eindimensional-zweifaktorielles Konstrukt* (didaktisch-methodisches [DM] und fachliches Wissen [F] als zwei Faktoren der *inhaltlichen* Dimension) Dabei laden die Items jeweils auf einem der Faktoren der inhaltlichen Dimension (entweder auf didaktisch-methodischem *oder* auf fachlichem Wissen).

*Modell 3* (s. Abb. 5, S. 42): *eindimensional-dreifaktorielles Konstrukt* (analytische Kompetenz [AK], Synthetische Kompetenz [SK] und Prozesskompetenz [PK] als drei Faktoren („Stufen“) der *formalen* Dimension)

Bei diesem Modell laden die Items jeweils auf einem der drei Faktoren der formalen Dimension (Komplexität der Informationsverarbeitung).

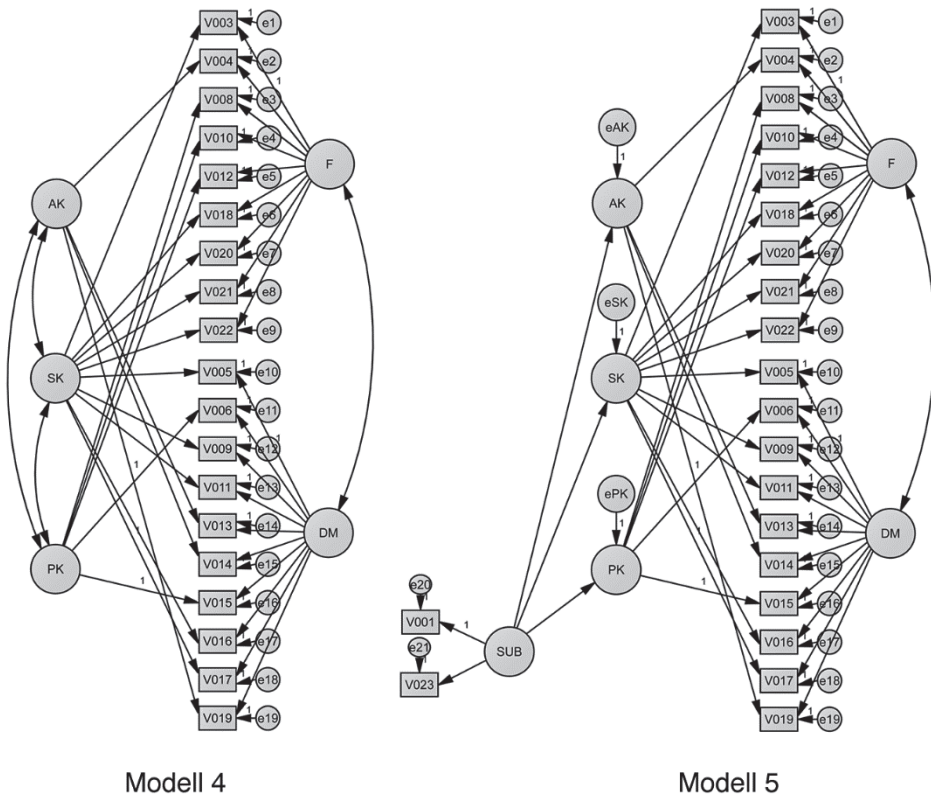


**Abb. 5:** Modell 2 und Modell 3 (aus: Plöger et al., 2015, S. 174)

Während in den Modellen 2 und 3 *keine* Doppelladungen auf der inhaltlichen und formalen Dimension zulässig waren, werden solche Doppelladungen in den Modellen 4 und 5 berücksichtigt.

*Modell 4* (s. Abb. 6, S. 43): *zweidimensionales Konstrukt* (Integration der inhaltlichen Dimension [mit zwei Faktoren] und der formalen Dimension [mit drei Faktoren])

Modell 4 kombiniert die Modelle 2 und 3 und stellt somit ein zweidimensionales Konstrukt dar, mit dem „die Varianz der manifesten Variablen durch die Faktoren der beiden unabhängigen Dimensionen erklärt werden“ (s. Plöger et al., 2015, S. 174) soll. Dabei erfasst jedes Item einerseits die *inhaltliche* Qualität des Wissens (didaktisch-methodisch *oder* fachliches Wissen) und andererseits dessen *formale* Qualität (Komplexitätsgrad der zu verarbeitenden Informationen).



**Abb. 6:** Modell 4 und Modell 5 (aus: Plöger et al., 2015, S. 175)

Es gilt aber die Restriktion, dass innerhalb der inhaltlichen Dimension (didaktisch-methodisches *und* fachliches Wissen) keine Doppelladungen erlaubt sind. Jedes Item soll *schwerpunktmäßig* entweder das didaktisch-methodische Wissen *oder* das fachliche Wissen erfassen. Dadurch können die Items dem entsprechenden Faktor der inhaltlichen Dimension zugeordnet werden. „Da aber davon auszugehen ist, dass die von jedem Item

erzeugte Varianz einen Rest des anderen Faktors mit abdeckt, wird diesem Sachverhalt durch die Korrelation von didaktisch-methodischem und fachlichem Wissen Rechnung getragen.“ (ebd., S. 175)

*Modell 5* (s. Abb. 6, S. 43): *zweidimensionales Konstrukt*

In Modell 5 wird, in Ergänzung zu Modell 4, das sogenannte *Subitizing*, also die Fähigkeit von ExpertInnen, Situationen schnell erfassen und sicher beurteilen zu können (s. Bromme, 1992), im Zusammenhang mit der formalen Dimension als *exogene Variable* herangezogen. Dazu wird zu Beginn und am Ende des Tests das identische Item eingesetzt, um mit dem globalen Urteil über die Unterrichtsstunde die „Varianz in den drei Faktoren der Komplexität der Informationsverarbeitung“ (Plöger et al., 2015, S. 175) aufzuklären.<sup>27</sup>

Hinsichtlich dieser fünf Modelle waren die durchzuführenden CFA an zwei zu prüfende Hypothesen geknüpft (s. Plöger et al., 2015, S. 175f.):

- *Hypothese 1*: „Die Modelle 1, 2 und 3 stellen keine adäquate Modellierung des Konstrukts Analysekompetenz dar.“ (ebd., S. 175)
- *Hypothese 2*: „Die Modelle 4 und 5 bilden das Konstrukt Analysekompetenz angemessen ab.“ (ebd., S. 176)

Um die Faktorenanalysen durchzuführen, wurde den einzelnen Faktoren der zu bildenden Modelle Testitems als Messindikatoren zugeordnet „unter der Voraussetzung, dass die zur Lösung der betreffenden Items notwendigen Anforderungen den theoretischen Annahmen entsprechen.“ (ebd., S. 173). Tabelle 2 stellt die Aufteilung der Items auf die jeweiligen Faktoren dar.

---

<sup>27</sup> Der Subitizing-Faktor wird über zwei Items zur *globalen* Bewertung der Unterrichtsstunde operationalisiert: „Das erste und letzte Item des Tests messen die Fähigkeit des Subitizing auf einer fünfstufigen Likertskala ...“ (Plöger et al., 2015, S. 177). Beide Items sind identisch formuliert („Wie bewerten Sie diese Unterrichtsstunde insgesamt?“) und zielen darauf ab, ob das rasche Urteil zu Beginn des Fragebogens über die Güte der zu analysierenden Unterrichtsstunde korrekt ist *und* mit der Bewertung der selbigen am Ende des Tests übereinstimmt (s. Plöger & Scholl, 2014a, S. 94).

**Tab. 2:** Zuordnung der Items zur formalen und inhaltlichen Dimension  
(aus: Plöger et al., 2015, S. 178)

		Formale Dimension		
		Analytische Kompetenz (AK, Stufe I)	Synthetische Kompetenz (SK, Stufe II)	Prozesskompetenz (PK, Stufe III)
<b>Inhaltliche Dimension</b>	Fachwissen (F)	V004	V003 V018 V020 V021 V022	V008 V010 V012
		Didaktisch-methodisches Wissen (DM)	V013 V014 V019	V006 V015
			V011 V016 V017	

Zur Veranschaulichung der Zuordnung der Items (Tab. 2) möchte ich vier Beispiele vorstellen, anhand derer sowohl die unterschiedliche Schwerpunktsetzung (didaktisch-methodisches Wissen [DM] oder fachliches Wissen [F]) als auch der unterschiedliche Komplexitätsgrad an Informationsverarbeitung deutlich wird (Analytische Kompetenz [AK] oder Synthetische Kompetenz [SK] oder Prozesskompetenz [PK]). Dabei greife ich auf Beispiele und die entsprechenden Erklärungen von Plöger et al. (2015) zurück, die in der Zeitschrift *Unterrichtswissenschaft* bereits veröffentlicht wurden (s. Plöger et al., 2015, S. 178ff.):

„Item V013 (DM; Stufe I, 1)

Verschafft sich der Lehrer einen Überblick darüber, ob seine Schülerinnen und Schüler dem Unterricht folgen können, oder tut er das nicht?

Dieses Item bezieht sich auf das didaktisch-methodische Prozessmerkmal Ergebnissicherung/Erfolgskontrolle. Zu seiner Lösung müssen keine komplexen Informationen verarbeitet werden, da es von den Probandinnen und Probanden lediglich verlangt wird, einzelne Situationen bzw. Handlungen des Unterrichts zu kategorisieren; sie können sich also an sichtbare Unterrichtsereignisse halten. So stellt die Lehrerin bzw. der Lehrer beispielsweise keinerlei (Kontroll- bzw. Rück-)Fragen, lässt wichtige Sachverhalte nicht (noch einmal) von den Schülerinnen und Schülern erklären bzw. wiederholen, ruft keine Zusammenfassungen von Teilergebnissen ab usw.“ (Plöger et al., 2015, S. 178)

„Item V004 (F; Stufe I, 2)

Im naturwissenschaftlichen Unterricht ist das Experimentieren eine wichtige methodische Kompetenz, die die Planung, Durchführung und Auswertung von Experimenten umfasst. Fördert dieser Unterricht diese methodische Kompetenz des Experimentierens oder nicht?

In der Formulierung des Items ist der Verlauf eines Experimentierprozesses im Fach Physik durch Angabe der einzelnen zu durchlaufenden Phasen vorgegeben. Die Lösung des Items ist durch Registrierung sichtbarer Ereignisse und ihrer graduellen Ausprägung (unklare Problemstellung, Durchführung des Experiments nach vorgegebenem Rezept, keine Diskussion der Messergebnisse usw.) möglich.“ (ebd., S. 178)



„Item V016 (DM; Stufe II, 3)

Wie geht die Lehrperson mit den einzelnen Antworten der Schülerinnen und Schüler um?  
Die Formulierung des Items setzt die Handlungen der Schülerinnen und Schüler (Antworten) mit denen der Lehrperson (Umgang mit den Antworten) in Beziehung. Fast durchgehend bleiben dabei Möglichkeiten der kognitiven Aktivierung ungenutzt. So nimmt die Lehrerin bzw. der Lehrer falsche Aussagen nicht zum Anlass einer konstruktiven Klärung. Sie oder er nutzt die Beiträge der Schülerinnen und Schüler, wenn überhaupt, sehr selektiv, ohne dieses Vorgehen verständlich zu machen. Zudem kommentiert sie bzw. er die Beiträge selbst, ohne andere einzubeziehen. Auf diese Weise entstehen keine Lerneffekte.“ (ebd., S. 179)

„Item V022 (F; Stufe II, 4)

Zum Ende der Stunde demonstriert der Lehrer als praktische Anwendung eine Fresnellinse. Wird den Schülerinnen und Schülern die physikalische Funktion dieser Linse verständlich oder bleibt sie ihnen unverständlich?

Zur Lösung dieses Items ist ein hinreichend hohes fachliches Vorwissen erforderlich: Die der Demonstration der Linse vorangehende Erarbeitung des Brechungsgesetzes hebt vorwiegend auf die Frage des Verhältnisses von Einfallswinkel und Brechungswinkel ab; die Funktion der Fresnellinse resultiert aber in erster Linie aus dem Sachverhalt, dass ein erheblicher Teil des Linsenmaterials eingespart werden kann, weil die senkrecht auf der Oberfläche einfallenden Lichtstrahlen keine Brechung erfahren, sondern erst bei Austritt aus der Linse (Übergang von Glas in Luft). Für einfallende Strahlen ist der Winkel also konstant = 0, für die austretenden gebrochenen Strahlen variiert er stark. Abgesehen davon müssen die Probandinnen und Probanden hier zwei Unterrichtsphasen, die vorangegangene Erarbeitung des Brechungsgesetzes und den intendierten Transfer, in Beziehung zueinander setzen; diese Leistung entspricht dem Komplexitätsgrad von Stufe 4.“ (ebd., S. 179)

1.6.2 Über CFA ermittelte faktorielle Struktur von Analysekompetenz

Die für die Prüfung der fünf Modelle relevanten Ergebnisse sind in Tabelle 3 dargestellt.<sup>28</sup>

Tab. 3: Fit-Indizes der fünf Modelle (aus: Plöger et al., 2015, S. 180)

Modell	$\chi^2$	df	P	$\chi^2/\text{df}$	CFI	RMSEA
1	657,825	152	,000	4,328	,810	,065
2	657,099	151	,000	4,352	,809	,065
3	652,145	149	,000	4,196	,821	,063
4	239,299	129	,000	1,855	,958	,033
5	302,646	166	,000	1,823	,960	,032

Die Fit-Indizes in Tabelle 3 belegen, dass die Modelle 1 bis 3 nicht geeignet sind, um das Konstrukt Analysekompetenz hinreichend zu erklären. Erst die Modelle 4 und 5 weisen einen Fit-Wert (CFI-Wert zwischen .950 und .969) auf, der als akzeptabel<sup>29</sup> gilt (s. z.B.

28 Zur Prüfung der vorgestellten fünf Modelle über die konfirmatorischen Faktorenanalysen wurde die Software Amos 21.0.0 (2013) verwendet.

29 Von einem guten Fit-Wert wird dann gesprochen, wenn der „an den Freiheitsgraden normierte Chi-Quadrat-Wert  $\chi^2/\text{df}$  zwischen 0 und 2, der RMSEA zwischen 0 und .05 und der CFI zwischen .970 und 1.00“ (Plöger et al., 2015, S. 180) liegt.



Moosbrugger & Schermelleh-Engel, 2012, S. 337f.) und somit das Konstrukt Analysekompetenz angemessen beschreibt.

Die Werte in Tabelle 4 zeigen „die für Modell 5 maßgeblichen standardisierten Ladungen und Korrelationen auf latenter Ebene“ (Plöger et al., 2015, S. 180). Der als exogene Variable herangezogene *Subitizing-Faktor* (SUB) zeigt, dass er eine hohe prädikative Kraft besitzt, um die Analysekompetenz auf den drei formalen Faktoren vorauszusagen.

**Tab. 4:** Standardisierte Ladungen und Korrelationen auf latenter Ebene in Modell 5 (aus: Plöger et al., 2015, S. 180)

	Standardisierter Schätzer	S.E.	C.R.	P
SUB → AK	,989	,057	14,346	,000
SUB → SK	,969	,058	14,465	,000
SUB → PK	,974	,035	11,960	,000
F ↔ DM	,854	,014	5,948	,000

Darüber hinaus besteht eine hohe Korrelationen zwischen den Faktoren der inhaltlichen Dimension (.85), die so interpretiert werden kann, dass eine Person, die im Test einen guten Wert hinsichtlich der didaktisch-methodischen Items erzielt hat, auch bei den fachlichen (bzw. fachdidaktischen) Aufgaben einen guten Wert aufweist, sodass angenommen werden kann, dass beide Wissensbereiche hochgradig vernetzt bei der Analyse von Unterricht genutzt werden (s. Gess-Newsome, 1999; Plöger et al., 2015, S. 180).

Zusammenfassend lassen sich die Ergebnisse dieses quantitativen Validierungsversuches folgendermaßen formulieren:

- 1) Die eindimensionale Modellierung der Analysekompetenz im Sinne eines G-Faktor-Modells muss im Hinblick auf die empirischen Daten abgelehnt werden.
- 2) Die theoretisch hergeleiteten Dimensionen, mit denen die inhaltliche und formale Qualität des Wissens abgebildet werden, zeigen, „dass weder die eine noch die andere Dimension für sich allein das Konstrukt Analysekompetenz angemessen abbilden kann“ (Plöger et al., 2015, S. 181).
- 3) Gute Fit-Werte werden erst dann erzielt, wenn beide Dimensionen *gemeinsam* in die Modellprüfung einbezogen wurden (s. Modell 4 und 5). Das Konstrukt Analysekompetenz kann somit als *zweidimensionales* Konstrukt interpretiert werden, welches sich aus einer inhaltlichen und einer formalen Dimension zusammensetzt. Oder um es abschließend mit den Worten von Plöger et al. (2015) zu konstatieren: „Um erfolgreich in der Analyse von Unterricht zu sein, muss man nicht nur über das richtige inhaltliche Wissen verfügen; zur Verarbeitung komplexer Informationen muss es auch hinreichend vernetzt bzw. hierarchisch organisiert sein.“ (ebd.)

### 1.6.3 Relevanz der Ergebnisse für den durchzuführenden qualitativen Validierungsversuch

Die Ergebnisse der von Plöger et al. (2015) durchgeführten CFA sprechen für die Validität des Konstrukts Analysekompetenz und stellen die Referenz für den von mir durchgeführten qualitativen Validierungsversuch dar. Mit Blick auf die erzielten Fit-Werte (s. Tab. 3, S. 46) hatte ich die Option, zwischen Modell 4 und 5 zu wählen, denn beide Modelle stellen adäquate Repräsentationen des Konstrukts dar. Meine Entscheidung ist letztlich zugunsten des fünften Modells ausgefallen.

Die Art der Erfassung des Subitizing-Faktors unterscheidet sich in den beiden Validierungsversuchen allerdings voneinander:

- 1) Im Validierungsversuch von Plöger et al. (2015) wurde den ProbandInnenen zunächst das gesamte Video ohne Unterbrechung gezeigt. Erst danach wurde der Fragebogen ausgefüllt. Dadurch konnten die ProbandInnen schon vor der Bearbeitung des Fragebogens einen Gesamteindruck von der Qualität der gesehenen Stunde gewinnen und diesen Eindruck dann über die Beantwortung des ersten Testitems, durch das zusammen mit dem letzten Item die Fähigkeit zum subitizing erhoben wurde (s. Plöger et al., 2015), wiedergeben.
- 2) In der von mir durchgeführten Studie wird das Video dagegen in fünf Abschnitte unterteilt (s. Kap. 2.3.3, S. 55) und durch entsprechende Interviewphasen „unterbrochen“. Dadurch können die ProbandInnen zu Beginn des Interviews nur (vage) Vermutungen über die Qualität des weiteren Unterrichtsverlaufs anstellen.

Trotz der unterschiedlichen Vorgehensweisen zur Messung entspricht aber auch die von mir verwendete Variante Plögers, Scholls und Seiferts Verständnis von subitizing. Bezogen auf die Analyse von Unterricht bedeutet *subitizing* nämlich, dass ExpertInnen schon nach kurzer Zeit eine präzise (wenn auch bis auf Weiteres hypothetische) Vorstellung von der Qualität der gesamten Stunde entwickeln, obwohl sie den weiteren Verlauf des Unterrichts noch gar nicht kennen. Durch die sukzessive Präsentation von Informationen in meiner Studie wird also ebenfalls die Fähigkeit operationalisiert, sich schnell ein hypothetisches Bild vom gesamten Unterrichtsprozess und dessen Lernwirksamkeit machen zu können.

## 2. Methodisches Vorgehen

### 2.1 Qualitative Inhaltsanalyse als methodischer Zugriff

Im vorangegangenen Kapitel habe ich die Zielstellung meiner Arbeit erläutert: Sie geht im Wesentlichen der Frage nach, ob sich die faktorielle Struktur des Konstrukts Analysekompetenz (wie in den Arbeiten von Plöger und Scholl (2014a) theoretisch modelliert und von Plöger et al. [2015] mittels CFA belegt) auf qualitativem Wege replizieren lässt. Konkreter formuliert: Lässt sich durch die Auswertung von Interviews (mit ExpertInnen und NovizInnen) nachweisen, dass eine angemessene Analyse von Unterricht nur dann gelingt, wenn das Wissen der betreffenden Personen eine hohe *inhaltliche* und *formale* Qualität besitzt? Dies wäre dann der Fall, wenn die Personen zum einen über ein hinreichendes didaktisch-methodisches und fachliches (incl. fachdidaktisches) Wissen und zum anderen über einen entsprechend hohen Grad der Komplexität der Informationsverarbeitung verfügen.

Mit Blick auf diese erkenntnisleitende Fragestellung sind zwei aufeinander abgestimmte Schritte zu gehen: Der 1. Schritt bestand darin, die theoretischen Annahmen, auf denen das Konstrukt Analysekompetenz basiert, vorzustellen (s. Kap. 1). In diesem Kapitel folgt nun der 2. Schritt, die Begründung der Wahl einer geeigneten qualitativen Methode zur Überprüfung des Konstrukts Analysekompetenz.

Diese Methode müsste mindestens zwei grundlegende Anforderungen erfüllen:

- 1) Um die faktorielle Struktur des Konstrukts Analysekompetenz abbilden zu können, müsste die betreffende Methode im Prinzip genau das leisten, was – wenn auch auf anderem Wege – eine Konfirmatorische Faktorenanalyse (CFA) ebenfalls leistet. Bei einer CFA wird eine große Menge manifester Variablen (Daten) auf eine relativ geringe Anzahl latenter Variablen (Konstrukte) reduziert. Die hier beabsichtigte Analyse von Interviews stellt einen analogen Fall dar: Es liegt umfangreiches manifestes Material (transkribierte Interviews) vor, das auf eine ihm zugrunde liegende latente Struktur (hier die faktorielle Struktur des Konstrukts Analysekompetenz) hin untersucht werden soll.<sup>30</sup>
- 2) Die Reduktion der Datenmenge ist eine wichtige Voraussetzung für das eigentliche Ziel einer CFA: Letztlich soll immer die Hypothese *geprüft* werden, ob die aus den ermittelten empirischen Daten gebildete Kovarianzmatrix die theoretischen Annahmen des zugrunde gelegten Konstrukts (so etwa das theoretische Konstrukt Analysekompetenz) abbilden kann.

Mit Blick auf diese beiden Anforderungen (Datenreduktion und Hypothesenprüfung) habe ich mich für die *Qualitative Inhaltsanalyse*<sup>31</sup>, wie sie insbesondere von Philipp Mayring

---

30 Mithilfe der Qualitativen Inhaltsanalyse erscheint es ebenfalls möglich, „die manifesten und latenten Inhalte des Materials in ihrem sozialen Kontext und Bedeutungsfeld zu interpretieren“ (Bortz & Döring, 2006, S. 329), um dadurch die Sichtweise der ProbandInnen herauszuarbeiten (s. Mayring & Fenzl, 2014).

31 In der Literatur existieren eine Vielzahl von Beschreibungen, was denn unter Qualitativer Inhaltsanalyse zu verstehen sei (s. z.B. Gläser & Laudel, 2010; Hsieh & Shannon, 2005; Kuckartz, 2016; Mayring, 2015). Mit

(2002, 2008, 2010, 2014, 2015) vertreten wird, entschieden. Diese Methode wird in der Regel zwar zur *Generierung* von Hypothesen eingesetzt (s. z.B. Ecarius & Miethe, 2011; Flick, 2009; Kelle, 2008; Kuckartz, 2016; Lamnek, 2010), aber sie stellt auch, wie ich im Folgenden zeigen möchte, ein effektives Mittel zur Überprüfung meiner Hypothese auf qualitativem Wege dar.

Was unter Qualitativer Inhaltsanalyse<sup>32</sup> zu verstehen ist, bringt Mayring in einer kurzen Formel auf den Punkt: „Qualitative Inhaltsanalyse will Texte systematisch analysieren, indem sie das Material schrittweise mit theoriegeleitet am Material entwickelten Kategoriensystemen bearbeitet.“ (Mayring, 2002, S. 114)

Diese Definition lässt sich sinngemäß etwa so paraphrasieren: Grundlegende Intention einer Qualitativen Inhaltsanalyse ist es, „Texte“ zu analysieren, wobei mit „Texten“ *Kommunikationsinhalte* jeglicher Art gemeint sind (Mayring, 1995, 2002, 2009, 2015; Mayring & Brunner, 2010), also nicht nur Texte im herkömmlichen Sinne, sondern z.B. auch Bilder, Plastiken, Filme usw. Eine entscheidende Voraussetzung für die systematische Analyse solcher Texte besteht darin, dass sie in *fixierter* Form vorliegen.

Das hier zu analysierende, umfangreiche und komplexe Material besteht aus schriftlich fixierten, d.h. transkribierten Interviews (jeweils neun ExpertInnen- und neun NovizInneninterviews). Ihre *systematische* Analyse soll Ordnung in diese Texte bringen, sie also *strukturell* verdichten, auf wesentliche Aussagen reduzieren und somit „auf den Begriff bringen“. Diese Begriffe werden in der Sprache der Qualitativen Inhaltsanalyse als *Kategorien* bezeichnet (Mayring, 2015, S. 51). Insofern ist eine Qualitative Inhaltsanalyse – analog zur Anwendung einer CFA – ein datenreduzierendes Verfahren.

Um dem Umfang und der Komplexität des Materials gerecht werden zu können, bedarf es in der Regel einer hinreichenden Zahl an Kategorien, die im Sinne der jeweils leitenden *theoretischen Fragestellung* in einen stringenten Zusammenhang gebracht werden müssen. Insofern bilden die einzelnen Kategorien ein *Kategoriensystem*, das nach Mayring das „zentrale Instrument der Analyse“ (Mayring, 2015, S.51) darstellt.

Im Falle der vorliegenden Arbeit besteht die Aufgabe demnach darin, ein Kategoriensystem zur Anwendung zu bringen, das die Analyse bzw. Interpretation der Interviews im

---

Blick auf diese Vielfalt kommt Schreier (2014) zu dem Schluss, dass es „die“ Qualitative Inhaltsanalyse nicht gibt (Schreier, 2014, Abs. 4) und „bisher eine allgemein akzeptierte Definition dessen, was den Kern qualitativ-inhaltsanalytischen Arbeitens ausmacht“ (Stamann, Janssen, & Schreier, 2016, Abs. 24), fehlt. Um dieses Desiderat zu beheben, erarbeiten Stamann et al. (2016) einen allgemeinen Kern qualitativ-inhaltsanalytischen Arbeitens: „Unserem Verständnis nach ist allen qualitativ-inhaltsanalytischen Verfahren die Systematisierung von Kommunikationsinhalten mit dem Ziel einer in hohem Maße regelgeleiteten Interpretation gemein. Daher handelt es sich bei der Qualitativen Inhaltsanalyse um eine Forschungsmethode zur Systematisierung von manifesten und latenten Kommunikationsinhalten. Die Methode zeichnet sich durch eine Vielzahl von forschungskontextuell spezifischen Verfahren aus. Als grundlegendes Instrument für die angestrebte Systematisierung der Inhalte dienen Kategoriensysteme.“ (Stamann et al., 2016, Abs. 9)

32 Seit einigen Jahren wird empfohlen, anstelle von „Qualitativer Inhaltsanalyse“ den Begriff der „qualitativ orientierten kategoriengeleiteten Textanalyse“ (Mayring, 2010, S. 604; Mayring & Fenzl, 2014, S. 544) zu nutzen. Denn dieser Begriff sei weniger problematisch und bringe deutlicher die *Grundintentionen* und das prinzipielle *Vorgehen* der Qualitativen Inhaltsanalyse zum Ausdruck. Die Qualitative Inhaltsanalyse laufe generell in zwei Schritten ab: In einem ersten *qualitativ-interpretativen* Akt werden Kategorien, die entweder induktiv oder deduktiv gebildet wurden, einem Text zugeordnet. Der zweite Schritt besteht dann darin, zu schauen, „ob bestimmte Kategorien mehrfach Textstellen zugeordnet werden können“ (Mayring & Fenzl, 2014, S. 544), um so auch *Quantifizierungen* vornehmen zu können.

Hinblick auf die faktorielle Struktur des Konstrukts Analysekompetenz ermöglicht. Dieses zugrunde liegende Erkenntnisinteresse sichert ein *selektives Vorgehen*, durch das spezifische, für das Konstrukt relevante Merkmale zur Sprache gebracht werden können. In diesem Sinne nutze ich die Qualitative Inhaltsanalyse zur *Überprüfung meiner Hypothese*, ob sich die faktorielle Struktur des Konstrukts Analysekompetenz auch auf qualitativem Wege replizieren lässt.

Dieses grundsätzliche Vorgehen, durch das Textstellen geeignete Kategorien zugeordnet werden, sei an einem konkreten Beispiel verdeutlicht. Die Faktoren der inhaltlichen und formalen Dimension sind in der hier vorliegenden Arbeit durch entsprechende Kategorien repräsentiert. So wird etwa innerhalb der inhaltlichen Dimension der Faktor *fachliches* (incl. *fachdidaktisches*) *Wissen* unter anderem durch die Kategorie *Experimentieren als zentrale Fachmethode* abgebildet (s. dazu im Detail Kap. 3.2.3, S. 104), wobei diese Kategorie folgendermaßen definiert ist:

Definition der Kategorie „Experimentieren als zentrale Fachmethode“:

Diese Kategorie beschreibt die Methode des Experimentierens als wechselseitigen Prozess von Theorie und Empirie. Dieser Prozess umfasst mehr als die bloße praktische Ausführung des Experiments selbst; denn diese ist eingebettet in theoretische Vorüberlegungen, Planung und Dateninterpretationen, stellt also nur einen Teil dieses Gesamtprozesses dar. (s. Kap. 3.2.4, S. 113)

Da die Definitionen solcher Kategorien in aller Regel sehr knapp und pointiert gefasst sind, werden sie zusätzlich durch erläuternde Beschreibungen ergänzt und ausdifferenziert.<sup>33</sup> Für die Kategorie *Experimentieren als zentrale Fachmethode* bedeutet das:

Beschreibung der Kategorie „Experimentieren als zentrale Fachmethode“:

Experimentieren ist ein mehrstufiger Prozess, der durch spezifische Fragen und Problemstellungen in Gang kommt. Daraus entwickeln sich theoriegeleitete Hypothesen, deren Überprüfung ein begründetes Versuchsdesign erfordert. Die Ausführung des Experiments selbst und die dadurch gewonnenen Messdaten lassen dann Rückschlüsse auf die Angemessenheit der Hypothesen zu. Dieser Prozess spiegelt sich im Unterricht in einer stimmigen Folge von entsprechenden Phasen: klare Problemstellung, Aufstellen und argumentatives Begründen von Hypothesen, Planung, Durchführung und Auswertung des Experiments. Die über diesen Prozess angestrebte Qualität des physikalischen Wissens hängt in hohem Maße davon ab, dass die einzelnen Phasen auf SchülerInnenseite zu präzisen und kohärenten Lernleistungen führen (Verstehen des Problems, Identifizieren möglicher Variablen, Begründen von Hypothesen, Entwerfen eines geeigneten Experiments, Durchführen des Experiments mit Blick auf die erwarteten Variablenzusammenhänge, Auswerten der ermittelten Daten unter der leitenden Problem- bzw. Fragestellung). In summa zielen diese Teilleistungen darauf, dass auf Seiten der SchülerInnen zunächst ein gedanklicher Lösungsprozess initiiert wird, der darin mündet, ein tiefes Verständnis davon zu entwickeln, was sie mit dem jeweiligen Experiment erkennen wollen. (s. Kap. 3.2.4, S. 113)

Wenn diese Kategorie (Experimentieren als zentrale Fachmethode) zur Analyse des Materials geeignet sein soll, müssten sich Textstellen finden lassen, auf die diese Kategorie angewandt werden kann. Der folgende Ausschnitt aus dem Transkript eines Interviews

33 Dabei stellen die Definitionen der Kategorien eine entscheidende Voraussetzung für die Qualität der Kodierungen dar. Denn „nur wenn die vom Forscher in Form von Kategorien intendierten Konstrukte genau definiert und ggf. durch Ankerbeispiele verdeutlicht sind, können die Kodierer nach einer Schulung oder zumindest auf der Basis einer schriftlichen Kodieranweisung das Ausgangsmaterial präzise verarbeiten“ (Bortz & Döring, 2006, S. 330).

erfüllt diese Forderung, denn der Interviewte moniert die Art und Weise des „Experimentierens“: Die Lehrperson habe die SchülerInnen nicht hinreichend aktiviert, das eigentliche Problem zu erfassen und dann selbstständig ein geeignetes Experiment zu konstruieren und durchzuführen. Insofern komme kein echter Experimentierprozess zustande, vielmehr nehme der Lehrer den SchülerInnen die Arbeit ab, indem er die zentralen Elemente des Prozesses (Aufstellen von Hypothesen, Planung und Durchführung des Experiments und die Interpretation der Daten) mehr oder weniger vorgibt.

Transkriptausschnitt aus einem Interviewprotokoll:

Exp-NW3: <h> Wobei das Entscheidende tatsächlich gewesen wär' wenn Schüler das selber geäußert hätten, sie hätten gemerkt es haben tatsächlich alle dasselbe 'raus //ja genau// so äh im Grunde genommen macht er ja eine Stunde wo zwar experimentiert wird //ja//, das Experiment aber gar keine Stellung äh gar keinen Stellenwert hat und zwar weder inhaltlich also es //ja// dient nicht zur Genese des Ergebnisses //ja// sondern das steht fest //ja//, <h> es wird aber auch nicht genutzt um zu reflektieren über die Art und den Stellenwert des Experimentierens überhaupt, //ja// im physikalischen Unterricht, <h> und konsequenterweise hätte man müsste man ihm raten wenn man in seiner Logik bliebe, äh soll er darauf verzichten //ja, hmhm// auf das Experiment //ja//. <h> ähm das würde den Unterricht aber an dieser Stelle auch nicht besser machen, weil= die Schüleraktivität //ja// nochmal reduzieren würde. #00:24:04-0# (Exp-NW3; Zeile 365-376)

Was an dieser Kategorie *Experimentieren als zentrale Fachmethode* beispielhaft verdeutlicht werden sollte (Definition der Kategorie, erläuternde Beschreibung, Auffinden geeigneter Textstellen), gilt entsprechend auch für die Anwendung aller anderen Kategorien des jeweils zugrunde liegenden Kategoriensystems, um so die Analyse und Interpretation, d.h. die *Kodierung*, des kompletten Textmaterials vornehmen zu können.

## 2.2 Vorüberlegungen zur Datengenerierung

Eine Grundvoraussetzung für die Analyse von Texten mittels Kategorien liegt in der *Verfügbarkeit* entsprechenden Textmaterials. Im günstigen Fall, der aber eher die Ausnahme darstellt, kann solches Material bereits vorliegen. In der Regel muss dieses Material also erst im Sinne des erkenntnisleitenden Interesses generiert werden. Das gilt auch für meine Untersuchung, mit der ich prüfen möchte, ob sich die von Plöger et al. (2015) durchgeführte quantitative Validierung des Konstrukts Analysekompetenz auch auf qualitativem Wege replizieren lässt.

Damit ein Vergleich der beiden Validierungsversuche möglich wird, müssen für die Generierung entsprechenden Textmaterials *Parameter* angesetzt werden. Diese Parameter ergeben sich insbesondere aus den Rahmenbedingungen der vorausgegangenen quantitativen Validierung:

- Die Testung der Analysekompetenz bei Plöger und Scholl (2014a) folgte dem *Experten-Novizen-Paradigma*, wodurch entsprechende Leistungsunterschiede hinsichtlich der Analysekompetenz sichtbar werden sollten. Infolgedessen ist es im Hinblick auf mein Vorgehen wichtig, Texte generieren zu lassen, in denen sich *unterschiedliche Grade von Analysekompetenz* spiegeln.

- Die Analysekompetenz hängt im Sinne des von Plöger et al. (2015) formulierten Konstrukts vom didaktisch-methodischen Wissen einerseits und vom fachlichen (incl. fachdidaktischen) Wissen andererseits ab. Für eine angemessene Analyse von Unterricht wird also nicht nur fächerübergreifendes didaktisch-methodisches Wissen benötigt, sondern auch fachliches Wissen. Deshalb ist es erforderlich, Texte zu erzeugen, in denen *beide Arten von Wissen* hinreichend dokumentiert sind.

Die Berücksichtigung dieser beiden Parameter hat – dies ist im Folgenden noch näher zu begründen – entsprechende Konsequenzen für die Auswahl der Personen, die solche Texte generieren können: Es müssen Personen sein, die sich im *Grad ihrer Analysekompetenz* wie auch in der *Fachzugehörigkeit* hinreichend unterscheiden. Die erste Bedingung wird durch Einbeziehung von *NovizInnen* (Studierenden) und *ExpertInnen* (Fach- und SeminarleiterInnen) realisiert; die zweite ist gegeben, wenn Texte von Personen zur Verfügung stehen, die *unterschiedliche Unterrichtsfächer* abdecken.

Damit die Vergleichbarkeit der beiden Validierungsversuche gegeben ist, sollten sich die Aussagen der ProbandInnen (im Fragebogen und im Interview) auf den *identischen* Stimulus, also auf dieselbe Unterrichtsstunde, beziehen. Denn Unterrichtsstunden besitzen eine eigene Faktur, die jeweils an ein bestimmtes Fach und an ein bestimmtes Thema gebunden ist. Aufgrund dieser Faktur machen dann auch nur ganz bestimmte *Facetten* didaktisch-methodischen und fachlichen Wissens für die Analyse Sinn. Würde dagegen nicht die gleiche Unterrichtsstunde zur Analyse eingesetzt, kämen schwerpunktmäßig andere Aspekte der didaktisch-methodischen und fachlichen Wissensfacetten zum Tragen bzw. die ProbandInnen würden unterschiedliche Wissensfacetten aktivieren.

## 2.3 Datengenerierung mittels Interviews

Im vorangegangenen Teilkapitel wurden die Anforderungen expliziert, die an die Generierung der Daten zu stellen sind. In diesem Teilkapitel soll nun das Vorgehen zur Datengenerierung selbst dargestellt werden. Ein geeignetes Verfahren zur Datenerhebung sehe ich dabei in der Methode des fokussierten, leitfadengestützten *Interviews* (s. z.B. Flick, 2011a; Hopf, 2009; Misoch, 2015).

### 2.3.1 Das Interview als Datengenerierungsquelle

Das *Interview* ist ein häufig eingesetztes Instrument in der Erziehungswissenschaft und wird vor allem dann gewählt, wenn die Erfassung subjektiver Sichtweisen und Sinnkonstruktionen im Zentrum des Erkenntnisinteresses steht (s. Flick, 2011a; Hron, 1982; Reinders 2015; Zierer, Speck, & Moschner, 2013). Dies ist auch bei der Analyse von Unterricht der Fall, denn die interviewten Personen drücken durch ihre Aussagen ihre Meinungen aus und rekonstruieren somit *aus ihrer Sicht* didaktisch-methodische Sinneinheiten von Unterricht.

Im Kontext von Forschungsprojekten stellt das Interview eine Befragung dar, bei der planmäßig vorgegangen wird und mindestens zwei Personen miteinander interagieren (s. z.B. Scheuch, 1973; Schreier, 2013b; Reinders, 2015). Dabei werden durch verbale Stimuli, z.B. durch Fragen, „*verbale Reaktionen (Antworten) hervorgerufen: Dies geschieht in*



*konkreten sozialen Situationen und unterliegt gegenseitigen Erwartungen. Antworten beziehen sich auf erlebte und erinnerte soziale Ereignisse, stellen Meinungen und Bewertungen dar.“* (Atteslander, 2010, S. 109, kursiv im Original)

Diese *gesprächsfunktionale* Konzeption des Interviews geht mit einer generellen Asymmetrie einher, da „die eine Person aufgrund eines wissenschaftlichen Interesses das Interview führt, die andere ... als Auskunftsperson zur Erfüllung dieses Interesses bei[trägt]“ (Helfferich, 2014, S. 564, Anm. MK).<sup>34</sup> Wenn sich die Kommunikation dabei aber zunehmend asymmetrisch entwickelt, können Texte entstehen, in denen die subjektiven Sichtweisen und Sinnkonstruktionen nur bedingt bzw. bruchstückhaft oder verzerrt zum Ausdruck kommen. Um solche Entwicklungen zu vermeiden, ist sicherzustellen, dass die handelnden Personen auf ihren gemeinsamen Vorrat an Deutungswissen zurückgreifen können. Denn dann agieren sie als Co-Konstrukteure des Gesprächs auf „Augenhöhe“ (s. Bogner & Menz, 2009).

Diese Co-Konstruktionen können aber aufgrund bestimmter Bedingungen gefährdet sein. Im Falle meiner Studie könnte dies z.B. der Fall sein, wenn ExpertInnen von Personen interviewt werden, die von den betreffenden ExpertInnen als wenig kompetent angesehen werden. Und umgekehrt könnten NovizInnen stark eingeschüchtert sein, wenn sie die interviewende Person als völlig überlegen definieren und die Person sie dies auch möglicherweise spüren lässt.

Um ein gewisses Maß an symmetrischer Kommunikation zu sichern, wurden die zu analysierenden Interviews mit den Fach- und SeminarleiterInnen (*ExpertInneninterviews*<sup>35</sup>) daher von Prof. Dr. Wilfried Plöger (berufliche Erfahrung als *Physiklehrer*, Publikationen im Bereich Allgemeiner *Didaktik* und *naturwissenschaftlicher Bildung*) und die Studierendeninterviews von mir als Wissenschaftlichem Mitarbeiter, der sich nach abgeschlossenem Lehramtsstudium in der Promotionsphase befindet, durchgeführt.<sup>36</sup>

### 2.3.2 Fokussierte Interviews

Vor dem Hintergrund dieser eher allgemeinen kommunikationskonstituierenden Überlegungen zur Interviewsituation habe ich mich für die spezifische Form des *fokussierten Interviews* (s. z.B. Merton & Kendall, 1993; Merton, Fiske, & Kendall, 1956) entschieden,

34 Diese allgemeine Asymmetrie lässt sich durch weitere Dimensionen in der Beziehung von interviewenden und interviewten Personen ausdifferenzieren. Helfferich (2014) unterscheidet in diesem Zusammenhang drei weitere Dimensionen: *Dimension der Machtrelation und der Sicherheit* (unterschiedliche Machtpotentiale von interviewten Personen [verfügt über Wissen] und interviewenden Personen [Kontrolle über Interviewsituation]); *Dimension Fremdheit/gemeinsamer Erfahrungshintergrund* (z.B. Unterschiede zwischen interviewten und interviewenden Personen hinsichtlich des Alters, Geschlecht, sozialen Status) und *Berücksichtigung unterschiedlich ausgeprägter Diskurskulturen* (Anpassung der Fragen an Reflexions-, Kommunikations- und Verbalisierungskompetenzen der interviewten Personen) (s. Helfferich, 2014, S. 564f.).

35 Ein Interview kann immer dann als *ExpertInneninterview* bezeichnet werden, wenn die ProbandInnen durch eine spezielle Auswahl und ihren Status definiert sind und somit als spezifische Zielgruppe eine besondere Art von Wissen besitzen (s. Bogner, Littig & Menz, 2014; Helfferich, 2014). Dabei ist jeweils im Einzelfall zu prüfen, wer als ExpertIn behandelt werden sollte (s. Meuser & Nagel, 1991, 2009, 2010). Die hier vorgenommene Auswahl wird in Abschnitt 2.3.4 ausführlich begründet.

36 Insbesondere die Forderung von Mayring, die Qualitative Inhaltsanalyse müsse in einem Kommunikationsmodell verortet werden, wird hier berücksichtigt (s. Mayring, 2015). Denn erst durch die Bestimmung des Kontextes (Sender, Empfänger etc.) kann geklärt werden, wie „das zu analysierende Material“ (Mayring & Brunner, 2010, S. 325) entstanden und zu bewerten ist.



das „auf einen bestimmten Gesprächsgegenstand“ (Fuhs, 2007, S. 71) vorab konzentriert ist, der im vorliegenden Fall durch einen visuellen Reiz in Form einer Videodarbietung gegeben ist.<sup>37</sup> Dabei handelt es sich um das in Kapitel 1 vorgestellte Video, das eine Unterrichtsstunde in der 11. Klasse einer Fachoberschule im Fach Physik zum Thema „Sneliussches Brechungsgesetz“ zeigt (s. Kap. 1.5.1, S. 38).

Die Auswahl und der Einsatz des Videos wurden oben (s. Kap. 1.5.1, S. 38f.) bereits begründet und beschrieben: Die Unterrichtsstunde weist charakteristische Schwachpunkte von Physikunterricht im Sinne der IPN-Videostudien auf, sodass die ProbandInnen ausreichend Äußerungs- und Anknüpfungsmöglichkeiten zur Formulierung von Handlungsalternativen haben.<sup>38</sup>

### 2.3.3 Der Interviewleitfaden

Die Realisierung fokussierter Interviews stützt sich auf einen entsprechenden *Interviewleitfaden* (s. Anhang A, S. 214), der immer dann notwendig ist, „wenn einerseits subjektive Theorien und Formen des Alltagswissens zu rekonstruieren sind und so maximale Offenheit gewährleistet sein soll, und wenn andererseits von den Interviewenden Themen eingeführt werden sollen und so in den offenen Erzählraum strukturierend eingegriffen werden soll“ (Helfferich, 2011, S. 179).

Mit dieser Aussage fordert Helfferich die Balancierung zweier Prinzipien der Interview- und Gesprächsführung ein, nämlich (1) ein Minimum an formaler und inhaltlicher Strukturierung versus (2) hinreichender Offenheit.

- 1) Ein Interviewleitfaden sollte in *formaler* und *inhaltlicher* Hinsicht ein Mindestmaß an Strukturierung bieten (s. Helfferich, 2014; Merton & Kendall, 1993), durch das die Interviews nach entsprechenden Vorgaben ablaufen können (Begrüßung, Einführung etc.) und dadurch in gewisser Weise vergleichbar werden (s. Bortz & Döring, 2006). Diese Forderung wird im vorliegenden Fall auf zwei Wegen sichergestellt:
  - Das Video stellt eine reale Physikunterrichtsstunde dar, die eine bestimmte Faktur aufweist. Mit Blick auf die Gliederung und den Verlauf des gezeigten Unterrichts ist das Video in fünf Abschnitte unterteilt, sodass die Darbietung des Videos nach jedem Abschnitt unterbrochen werden kann. Diese Segmentierung stellt in formaler Hinsicht

37 Videos werden seit mehreren Jahrzehnten in der Aus-, Fort- und Weiterbildung von Lehrpersonen eingesetzt (s. z.B. Helmke & Helmke, 2004; Herzmann & Proske, 2014; Hugener, Rakoczy, Pauli, & Reusser, 2006; Kramer, König, Kaiser, Ligtvoet, & Blömeke, 2017; Krammer & Hugener, 2014; Krammer & Reusser, 2005; Pauli & Reusser, 2006), allerdings hat die Nutzung von videografiertem Unterricht vor allem in den letzten Jahren stark zugenommen (Seidel & Thiel, 2017). Seidel und Thiel (2017) führen diese Zunahme, neben dem technischen Fortschritt (z.B. höhere Aufnahmequalität, bessere Speicherungsmöglichkeiten), vor allem auf die *Qualitätsoffensive Lehrerbildung* zurück, in deren Rahmen an vielen Universitäten Videos „zur Optimierung von Lehrerbildung“ (Seidel & Thiel, 2017, S. 5) eingesetzt werden. Im Hinblick auf den vermehrten Einsatz von Videos identifizieren die beiden Forscherinnen drei Trends der videobasierten Lehr-Lernforschung, die in den kommenden Jahren den *evidenzbasierten* Einsatz von Videos in der Lehr-Lernforschung weiter forcieren könnten bzw. werden: „1. Optimierung der prozessbezogenen Erfassung des Lehr-Lern-Geschehens, 2. Effekte videobasierter Intervention zur Verbesserung der Lehrerbildung, 3. Entwicklung von videobasierten Tests in der Erfassung von Lehrkompetenzen“ (ebd.).

38 Das *Fokusinterview* wird verstanden als „ein Interviewverfahren, vor dessen Beginn eine von allen Befragten erlebte Stimulusituation (Film, Radiosendung, gelesener Text, erlebtes Ereignis, Experiment) steht. Das Interview ist darauf fokussiert, auszuleuchten, wie diese Situation subjektiv empfunden wurde und was davon wie wahrgenommen wurde.“ (Przyborski & Wohlrab-Sahr, 2014, S. 135)

also eine zeitliche Strukturierungshilfe dar, indem nach jedem Abschnitt das Gespräch mit den ProbandInnen durch entsprechende (offene) *Eröffnungsfragen* initiiert wird und Kommentare zum Gesehenen von den Interviewten erbeten werden.

- Darüber hinaus bildet das Thema der Unterrichtsstunde einen Gesamtrahmen, aus dem sich in *inhaltlicher* Hinsicht zahlreiche Aspekte ergeben, die im Interview zur Sprache kommen könnten, wie z.B. die Abfolge der Phasen, die damit zusammenhängenden Teilziele der Unterrichtsstunde, der Medieneinsatz, die Aktivierung der SchülerInnen, deren Fehlvorstellungen, der Experimentierprozess usw.
- 2) Um das Prinzip der *hinreichenden Offenheit* zu gewährleisten, sollen die interviewten Personen im Anschluss an die Betrachtung der Videoabschnitte möglichst unvoreingenommen ihre subjektiven Sichtweisen äußern und nicht theoretisch postulierte Kategorien abarbeiten, die durch den Leitfaden abgefragt würden (s. Hopf, 1978).<sup>39</sup> Dieses *Prinzip der Offenheit* gilt für ExpertInnen wie NovizInnen gleichermaßen:
  - ExpertInnen verfügen über umfangreiches und hinreichend vernetztes Wissen, sodass enge Fragestellungen unnötig sind und sogar als Gängelung empfunden werden könnten.
  - Bei NovizInnen ist das zur Analyse von Unterricht notwendige Wissen dagegen erst ansatzweise vorhanden. Von daher könnten auf bestimmte Aspekte eingeeengte Fragestellungen den interviewten Personen indirekte oder direkte Hinweise geben, durch die sie auf Sachverhalte aufmerksam würden, die sie ohne diese Hinweise selbstständig nicht wahrgenommen hätten. Insofern sind offene Fragen besser als geschlossene bzw. enggeführte geeignet, um die *Grenzen* des ProbandInnenwissens zu erkennen.

Um dem Anspruch der möglichst freien Äußerung gerecht zu werden und das „Spektrum der im Interview angeschnittenen Problemstellungen“ (Friebertshäuser & Langer, 2010, S. 354) nicht zu eng zu fassen, kamen unterschiedliche *Frageotypen* zum Einsatz. Im vorliegenden Fall wurde dies hauptsächlich durch das Stellen *unstrukturierter (offener)* Fragen realisiert (Beispiel: „Was haben Sie in diesem Unterrichtsausschnitt beobachten können?“). Solche offenen Fragen dienten zunächst als *Redeimpuls* und sollten die Befragten in einem ersten Schritt veranlassen, möglichst *freie* Antworten zu formulieren (s. Friebertshäuser & Langer, 2010; Merton & Kendall, 1993).

Darüber hinaus wurden, wenn das Gespräch nur schwerlich in Gang kam oder bestimmte thematische Aspekte nicht angesprochen wurden, *halbstrukturierte* Fragen verwendet, bei denen zwar ein bestimmter Aspekt der Unterrichtsanalyse angesprochen wurde, aber die Reaktion (die Antwort) der interviewten Person frei erfolgen konnte („Was sagen Sie zur Auswahl der Beispiele?“).<sup>40</sup>

39 Dabei soll betont werden, dass das Prinzip der Offenheit eine Art *Ideallinie* darstellt, „die Orientierung gibt, die aber nie vollständig einzulösen ist“ (Helfferich, 2014, S. 562). Denn einerseits ist das Interview eine soziale Situation, die durch die beteiligten Personen konstruiert wird (s. oben) und andererseits durch ein entsprechendes Forschungsinteresse motiviert ist, sodass eine vollkommene Offenheit nur bedingt zielführend sein kann (s. Helfferich, 2014).

40 Im Rahmen des fokussierten Interviews und der produktiven Gestaltung einer Interviewsituation haben Merton & Kendall (1993) vier Qualitätskriterien für den Ablauf eines fokussierten Interviews formuliert: *Nicht-Beeinflussung*, *Spezifität*, *Erfassung eines breiten Spektrums* sowie *Tiefgründigkeit* (ebd., S. 179ff.). Diese Kriterien waren auch in dieser Untersuchung handlungsleitend, denn die ProbandInnen sollten möglichst wenig beeinflusst werden und die Chance erhalten, eine individuelle Einschätzung und Stellungnahme hinsichtlich des Reizes (hier: videografierte Unterrichtsstunde) vornehmen zu können (*Nicht-Beeinflussung*).

Um den zunächst erstellten Leitfaden zu optimieren, wurden Probeinterviews mit Studierenden geführt. Dabei wurde insbesondere geprüft, wie das Prinzip der Offenheit eingehalten werden konnte. So konnten Fragen ausfindig gemacht und eliminiert werden, die nur bedingt eine offene Antwort evozierten, wie zum Beispiel: „Welche Effekte kann diese Vorgehensweise auf die Schüler haben?“ Fragen wie diese gaben den interviewten Personen Hinweise, nach Effekten zu suchen, die sie eigenständig eher nicht wahrgenommen hätten.

### 2.3.4 Auswahl der TeilnehmerInnen

Um im Folgenden die Auswahl der TeilnehmerInnen zu begründen, greife ich auf die weiter oben bereits genannten und begründeten Parameter (Grad der Analysekompetenz; Fachzugehörigkeit) zurück (s. S. 52f.). Im Sinne dieser beiden Parameter sind Personen einzubeziehen, die sich zum einen im Grad ihrer Analysekompetenz und zum anderen hinsichtlich ihrer Fachzugehörigkeit unterscheiden sollten. Im Detail heißt das:

- 1) Die Untersuchungen von Plöger und Scholl folgten dem *Experten-Novizen-Paradigma*, in dem davon ausgegangen wird, dass Leistungsunterschiede zwischen den TeilnehmerInnen (hier: hinsichtlich der Analysekompetenz von Unterricht) deutlich werden. Die Leistungsunterschiede beziehen sich auf den unterschiedlichen Grad der Analysekompetenz, also auf die Fähigkeit, komplexere und weniger komplexere Unterrichtssituationen (unter didaktisch-methodischen und fachlichen Gesichtspunkten) wahrnehmen zu können. Im Sampling sollten daher Personen vertreten sein, die einerseits Unterricht sehr gut analysieren können, und andererseits Personen, die (vermutlich) über weniger gute Analyseleistungen verfügen (s. Kap. 1.5.2, S. 40). Da sich mit Blick auf die Ausprägung der Analysekompetenz im Projekt von Plöger und Scholl gezeigt hatte, dass Studierende nur relativ geringe Analyseleistungen erbrachten, Fach- und SeminarleiterInnen dagegen sehr hohe (s. Plöger & Scholl, 2014a), habe ich für meine Studie Studierende als NovizInnen und Fach- und SeminarleiterInnen als ExpertInnen gewonnen. Bei der Rekrutierung dieser Personen (ExpertInnen und NovizInnen) wurde folgendermaßen vorgegangen:
  - Im Rahmen der vorausgegangenen quantitativen Untersuchung von Plöger und Scholl (2014a) wurden 200 Fach- und SeminarleiterInnen befragt. Aus diesem Personenkreis sollten diejenigen ProbandInnen für ein Interview gewonnen werden, die im Test eine besonders hohe Analyseleistung gezeigt und sich somit als ExpertInnen erwiesen hatten.<sup>41</sup> Deshalb wurden die Seminarleitungen gebeten, den Kontakt zu diesen Personen herzustellen, die daraufhin ihre Anonymität preisgaben und sich für ein Interview zur Verfügung stellten.

---

Des Weiteren sollten die interviewten Personen keine globalen Einschätzungen und Deutungen von Unterricht vornehmen, sondern konkrete, situationsspezifische (subjektive) Deutungen über den gezeigten Unterricht anbringen (*Spezifität*). Darüber hinaus wurde den ProbandInnen aber auch Raum für unerwartete Reaktionen gelassen, sodass sie nicht auf einzelne Reaktionen festgelegt wurden, sondern sich frei und umfassend äußern konnten (*Erfassung eines breiten Spektrums*). Diese Freiheit mündete darin, dass eine Tiefgründigkeit der Analyse angestrebt wurde und die interviewten Personen durchaus affektive, kognitive und wertbezogene Reaktionen zeigen konnten (*Tiefgründigkeit*).

<sup>41</sup> Die ExpertInnen, die ausgewählt wurden, haben sich im Test von Plöger und Scholl (2014a) als „Top-Score-Innen“ erwiesen, da sie beim Test am besten abschnitten, indem sie 20 von möglichen 23 Punkten erreichten.

- Die Studierenden (NovizInnen) konnten an der Universität zu Köln aus Seminaren im Studiengang Bildungswissenschaften (Veranstaltungen der Projektleiter) in genügender Zahl rekrutiert werden. Zu Beginn des Sommersemesters 2013 meldeten sich Studierende freiwillig, die als *Incentive* ein Geschenk im Wert von 20 Euro erhielten.
- 2) Die Fachzugehörigkeit bildet einen weiteren Parameter, der die Analysekompetenz der ProbandInnen bestimmt. So haben Plöger und Scholl (2013, 2014a) zeigen können, dass Personen, die das Fach Physik bzw. ein anderes naturwissenschaftliches Fach und/oder Mathematik studierten bzw. unterrichteten, bei der Analyse des eingesetzten Videos denjenigen Personen überlegen waren, die keines dieser Fächer unterrichteten. Vor dem Hintergrund dieses Ergebnisses sollten in meiner Studie Personen repräsentiert sein, die sich hinsichtlich der Fachzugehörigkeit unterscheiden. Auf diese Weise lässt sich die Hypothese testen, ob Personen, die das Fach Physik oder ein affines Fach unterrichten, bei der Analyse von Unterricht intensiver fachliches Wissen aktivieren als Personen, die nicht „vom Fach sind“. Deshalb wurde das Sampling in drei Gruppen ausdifferenziert: TeilnehmerInnen, die
  - das Fach Physik,
  - ein anderes naturwissenschaftliches Fach (ggf. in Kombination mit Mathematik) oder
  - keines dieser Fächer unterrichten bzw. studieren.

Unter Berücksichtigung dieser Differenzierung wurden insgesamt 18 Personen interviewt, die sich im Status (ExpertInnen – NovizInnen) und hinsichtlich der Fachzugehörigkeit entsprechend unterscheiden. Tabelle 5 zeigt die Zusammensetzung des Samplings (*qualitativer Stichprobenplan*<sup>42)</sup> im Detail<sup>43</sup>:

**Tab. 5:** Zusammensetzung des Samplings

Fach	ExpertInnen	NovizInnen
Physik (Ph)	3	3
mind. ein naturwissenschaftliches Fach (NW)	3	3
kein naturwissenschaftliches Fach (KNW)	3	3
<b>Total</b>	<b>9</b>	<b>9</b>

42 In der Forschungsliteratur wird dieses Verfahren auch als ein *non-probabilistisches Verfahren* der Stichprobenziehung bezeichnet, bei dem nicht die *Zufallsauswahl* der TeilnehmerInnen eine Rolle spielt, sondern die *bewusste bzw. absichtsvolle* Stichprobenziehung im Vordergrund steht (s. Schreier, 2011). Dieses Verfahren zeichnet sich dadurch aus, dass die Stichprobe „gezielt nach bestimmten Kriterien aus der Grundgesamtheit ausgewählt“ (Schreier, 2013a, S. 194) wird, mit der Absicht, den größtmöglichen Erkenntnisgewinn in Bezug auf die Fragestellung anzustreben (s. Schreier, 2010, S. 241). Nach Flick wird dieses Verfahren auch als *A-priori-Determination* bezeichnet (s. Flick, 2009, S. 155).

43 Die ausgewählten ExpertInnen (drei weibliche, sechs männliche Personen) weisen eine mindestens 15-jährige Unterrichtserfahrung als Lehrpersonen vor. Die Studierenden (fünf weibliche, vier männliche Personen) haben im Gegensatz dazu durchschnittlich 15 Stunden Unterrichtserfahrung im Rahmen von Praktika.

### 2.3.5 Durchführung der Interviews

Neben den bisherigen theoretischen Vorüberlegungen sollen nun die organisatorischen Vorbereitungsmaßnahmen sowie die (technischen) Abläufe der Interviews skizziert werden, um die Darstellung der Datengenerierung abzurunden.

Nachdem die ausgewählten ProbandInnen ihre Teilnahme signalisierten, wurden Räumlichkeiten ausgewählt, in denen die Interviews stattfinden konnten. Während die Experteninterviews mit den Fach- und SeminarleiterInnen in den jeweiligen Studienseminaren durchgeführt wurden, fanden die Studierendeninterviews in den Räumen der Universität zu Köln statt.

Zur Gewährleistung einer unverstellten Sicht auf das zu analysierende Video wurden entweder ein Beamer oder ein Computer eingesetzt, um das Video an eine Wand oder über einen (entsprechend großen) Bildschirm projizieren zu können. Bei der Gestaltung der Interviewsituation wurde Wert auf eine entspannte Gesprächsatmosphäre gelegt. Daher sollten sich die interviewte und die interviewende Person nicht frontal gegenüber sitzen (ähnlich einer Verhör-Situation), sondern beide sollten „über Eck“ miteinander kommunizieren.

Der Interviewverlauf wird aber nicht nur durch solche organisatorischen Vorkehrungen beeinflusst, sondern auch durch den (technischen) Ablauf des Interviews. Im Rahmen von Leitfadeninterviews wird ein vierphasiger Aufbau empfohlen. Die einzelnen Phasen bezeichnet Reinders als Einstiegsphase, Aufwärmphase, Hauptphase und Ausstiegsphase (Reinders, 2015, S. 99ff.), die in der hier durchgeführten Untersuchung berücksichtigt wurden.

- *Einstiegsphase*: Die ProbandInnen wurden zu Beginn über das *Procedere* (den Ablauf, die technische Aufzeichnung, die anonymisierte Weiterverarbeitung usw.) aufgeklärt. Alle TeilnehmerInnen stimmten in *mündlicher* Form der anonymisierten Aufzeichnung und Weiterverarbeitung der Interviewdaten zu.
- *Aufwärmphase*: Um das Gespräch inhaltlich „in Gang zu setzen und gleichzeitig praktisch in den Gesprächsmodus einzuführen“ (Reinders, 2015, S. 100), wurde versucht, den TeilnehmerInnen (insb. den Studierenden) ihre (mögliche) Unsicherheit vor dem eigentlichen Interviewbeginn zu nehmen. Es wurde ihnen erklärt, dass sie ein 15-minütiges Video einer Physikunterrichtsstunde zum Thema des Snelliusschen Brechungsgesetz sehen werden, das in fünf Abschnitte unterteilt ist, und dass sie nach jedem Stopp die Möglichkeit bekommen, sich frei zum Gesehenen äußern zu können.
- *Hauptphase*: In der zeitlich umfangreichsten Phase des Interviews wurden die im Leitfaden dargestellten Themen und Fragen zur Sprache gebracht, die aber nicht rigoros (im Sinne der „Leitfaden-Bürokratie“ [Hopf, 1978]) abgearbeitet wurden. Vielmehr blieb den interviewten Personen ausreichend Antwortspielraum, um den Unterricht möglichst unvoreingenommen, d.h. nach eigenen, selbstgewählten Aspekten analysieren zu können.<sup>44</sup>

---

44 Die interviewende Person hatte dabei die Aufgabe, den Aussagen der interviewten Person aufmerksam zu folgen, um mittels eines *neutral-weichen Interviewstils* weitschweifige Kommentare behutsam zum Thema zurückzulenken oder durch Nach- und Verständnisfragen (Beispiel: „Könnten Sie dies bitte genauer erläutern?“) die Aussagen der ProbandInnen spezifizieren zu lassen.

- *Ausstiegsphase*: Zum Abschluss des Gesprächs bekamen die TeilnehmerInnen die Möglichkeit, eine abschließende, zusammenfassende Beurteilung der Unterrichtsstunde vorzunehmen, mit der sie noch einmal die thematischen Aspekte ansprechen konnten, die aus ihrer Sicht eine besonders hohe Relevanz besaßen.

## 2.4 Datenaufbereitung – Transkription

Nach Aufzeichnung der Interviews mittels Tonbandgerät müssen die vorliegenden Audiodaten verschriftlicht werden. Dieser Prozess der Verschriftlichung verbaler (mündlicher) Daten (hier: Tonbandaufzeichnungen) in fixierte Kommunikation (schriftlich) wird *Transkription* genannt (s. Langer, 2010).<sup>45</sup>

Eine solche Transkription ist als *Transformationsprozess* zu verstehen, der kein vollständiges Abbild der Aufzeichnungen, sondern einen ersten interpretativen Akt darstellt (s. Fuhs, 2007), der „Dinge sichtbar oder unsichtbar machen“ (Langer, 2010, S. 524) kann. Um den Akt der Verschriftlichung „transparent zu machen“ und nachvollziehbar zeigen zu können, „wie das jeweilige Transkript entstanden ist“ (Dresing & Pehl, 2010, S. 723), werden Transkriptionsregeln und Notationszeichen benötigt, wobei die Transkriptionsregeln zusammen ein *Transkriptionssystem* bilden (s. ebd.).

In der Literatur finden sich eine ganze Reihe solcher Transkriptionssysteme, die sich in ihrer Detailliertheit und Genauigkeit unterscheiden (s. Dittmar, 2004; Kowal & O’Connell, 2009; Kuckartz, 2016; Langer, 2010). Die Existenz verschiedener Transkriptionssysteme deutet darauf hin, dass mit ihnen unterschiedliche Intentionen und Ansprüche verbunden sind, die auf die spezifische Fragestellung bzw. das leitende Erkenntnisinteresse abgestimmt sind. Mit Blick auf meine Fragestellung habe ich mich dafür entschieden, aus den verschiedenen Systemen die für meine Untersuchung wichtigen formalen Anforderungen und Transkriptionsregeln zu extrahieren. Als formale Anforderungen wurden berücksichtigt<sup>46</sup>:

- Es wurden Transkripte als Materialbasis benötigt, die nicht durch eine *vorschnelle Selektion* die inhaltlichen Aussagen der ProbandInnen einschränken oder gar verfälschen. Deshalb wurden alle Interviews *vollständig* transkribiert, um eine frühzeitige Reduktion des Datenmaterials zu verhindern und der Forderung von Strauss und Corbin (s. 1996, S. 15) gerecht zu werden, zunächst lieber mehr als zu wenige Textteile zu verschriftlichen.
- Die Abfolge der Redebeiträge wurde *chronologisch* in *Zeilenschreibweise* dokumentiert.

45 Prinzipiell werden zwei Ziele mit der Verschriftlichung der vorliegenden Audiodateien verfolgt, die vornehmlich darin bestehen, „dass sie [hier: die aufgezeichneten Interviews, Anm. MK] sowohl für Auswertungsverfahren genutzt als auch den Lesenden einer wissenschaftlichen Auswertung zugänglich gemacht werden können“ (Dresing & Pehl, 2010, S. 723).

46 Diese formalen Anforderungen dienen dazu, eine *Einheitlichkeit* bei der Transkription herzustellen, die dann die *Vergleichbarkeit* zwischen den Interviews bzw. Transkripten gewährleisten soll (s. Langer, 2010, S. 519).

- Jedes Transkript erhält einen *Transkriptionskopf* mit den allgemeinen Angaben zur Gesprächssituation (und Tonbandaufnahme) sowie mit anonymisierten Angaben zur interviewten Person (s. Flick 2009).<sup>47</sup>
- Des Weiteren sollten die einzelnen Äußerungen der ProbandInnen möglichst nah am gesprochenen Wort wiedergegeben werden. Insofern wurde die Transkription als *literarische Umschrift*<sup>48</sup> realisiert, die zusätzlich auch Überlappungen der Redebeiträge aufnimmt und darüber hinaus die *lautliche* Gestaltung des Redebeitrags (z.B. Lautstärke, Betonung, Dehnung) berücksichtigt sowie teilweise *non-* wie auch *paraverbale* Kommunikationsmerkmale, wie Lachen oder Hüsteln, erfasst.<sup>49</sup>

Über diese formalen Anforderungen hinaus orientiert sich der Transkriptionsprozess an einer Reihe von Regeln, die im Anhang vollständig wiedergegeben werden (s. Anhang B, S. 215 & Anhang C, S. 216). Exemplarisch seien hier einige genannt:

**Tab. 6:** Beispiele für verwendete Transkriptionsregeln

Zeichen	Bedeutung
.	deutlich fallende Intonation
;	leicht fallende Intonation
?	deutliche Frageintonation
(.)	kurze Sprechpause oder Verzögerung
(2)	lange Sprechpause mit Sekundenangabe
brau-	Wortabbruch
Wort	Betonung

Damit dieses Transkriptionssystem in einheitlicher Form zur Anwendung kommen kann, müssen die Transkribierenden (studentische Hilfskräfte), die dieses System nicht kennen, darin geschult werden. Dies geschah in folgender Weise: Zunächst wurden die Studierenden in die Intention und den Prozess der Transkription sowie in das Regelsystem eingeführt. Zur schrittweisen Einführung in dieses System waren entsprechende Übungen an beispielhaften Textstellen notwendig. Daran anschließend wurde an Beispieltranskripten die (flexible) Anwendung des gesamten Regelsystems getestet, bevor die Studierenden die Aufgabe erhielten, die ersten Minuten eines Interviews selbstständig zu transkribieren. Als technisches Hilfsmittel wurde dabei die (Transkriptions-)Software f4<sup>50</sup> verwendet. Nachdem die Studierenden hinreichende Sicherheit im Befolgen aller Regeln zeigten, wurden die Audiodateien von ihnen komplett verschriftlicht. Anschließend erfolgte ein Korrekturdurchgang (s. den Auszug aus einem Interviewtranskript, Anhang D, S. 217f.).

47 Weitere personenbezogene Angaben (z.B. Geschlecht, Berufserfahrung) wurden in einer zusätzlichen Datenmatrix festgehalten (s. Flick, 2009).

48 Die literarische Umschrift wird zu den Formen der wörtlichen Transkription gezählt (s. Mayring, 2002), bei der sprachliche Besonderheiten berücksichtigt werden (s. Gläser-Zikuda, 2015; Langer, 2010).

49 Neben der Erfassung der sprachlichen Besonderheiten wurden auch Zusatzinformationen, die über das Wortprotokoll hinausgehen, erfasst. Diese Variante wird auch als „kommentierte Transkription“ bezeichnet (s. Gläser-Zikuda, 2015; Mayring, 2002).

50 Dresing & Pehl entwickelten die Transkriptionssoftware f4, deren Stärke vor allem in ihrer intuitiven, gut handhabbaren Bedienung liegt (s. Dresing & Pehl, 2010).



## 2.5 Kategorienbasierte Kodierung der Transkripte

Die Transkripte stellen das Ausgangsmaterial dar, das im Verlauf der Qualitativen Inhaltsanalyse ausgewertet werden soll. Zu diesem Zweck kommen nun die bereits in Kapitel 2.1 erwähnten Kategorien zum Einsatz, mit denen das Textmaterial kodiert werden soll. Im Folgenden wird beschrieben, wie die für die Analyse notwendigen Kategorien gebildet wurden (s. Kap. 2.5.1). Anschließend erläutere ich den Vorgang der Kodierung des Materials (s. Kap. 2.5.2) und das dazu angewandte Konsensverfahren (s. Kap. 2.5.3). Abschließend wird die Prüfung der Reliabilität (s. Kap. 2.5.4) und Validität (s. Kap. 2.5.5) der vorgenommenen Kodierungen beschrieben.

### 2.5.1 Entwicklung eines Kategoriensystems

Für die Bildung von Kategorien, die den zentralen Ausgangspunkt der Materialanalyse darstellen, werden im Rahmen der Methodendiskussionen zwei prinzipielle Wege diskutiert.<sup>51</sup> Zum einen können Kategorien *induktiv* gebildet werden. Bei diesem Vorgehen werden Kategorien direkt aus dem Material heraus generiert, ohne dass man sich auf „vorab formulierte Theoriekonzepte“ (Mayring, 2015, S. 85) bezieht. Dieses induktive Vorgehen soll helfen, das Material „ohne Verzerrungen durch Vorannahmen des Forschers“ (Mayring, 2015, S. 86) in den Blick zu nehmen. Zum anderen können Kategorien *deduktiv* an das Material herangetragen werden. Damit ist gemeint, dass von bestehenden theoretischen (Vor-)Annahmen ausgegangen wird, um Kategorien zu definieren, die anschließend auf das Textmaterial angewandt werden.

Ob der eine oder der andere Weg der Kategorienbildung gewählt wird, „hängt in starkem Maße von der Forschungsfrage, der Zielsetzung der Forschung und dem Vorwissen ab, das bei den Forschenden über den Gegenstandsbereich der Forschung vorhanden ist“ (Kuckartz, 2016, S. 63). Ich habe das deduktive Verfahren gewählt, weil ich mich an den theoretischen Annahmen orientiere, aus denen heraus Plöger et al. (2015) das Konstrukt Analysekompetenz definieren und die ihrem quantitativen Validierungsversuch zugrunde lagen. Im Sinne dieser Annahmen ist das Konstrukt Analysekompetenz durch zwei Dimensionen gekennzeichnet: durch eine inhaltliche und eine formale. Die inhaltliche Dimension ist weiter durch zwei Faktoren differenziert (didaktisch-methodisches und fachliches Wissen). Die formale Dimension wird durch fünf Stufen (= Faktoren) der Komplexität der Informationsverarbeitung wiedergegeben. Die nachfolgende Tabelle 7 stellt überblicksartig dar, wie die zu entwickelnden Kategorien diesen Dimensionen bzw. Faktoren zugeordnet werden.

---

51 Diese beiden Varianten stellen mehr oder weniger Extrempole dar, die in der Forschungspraxis nicht unbedingt in Reinform vorkommen müssen, sondern auch als *Mischformen* möglich sind (s. Bortz & Döring, 2006; Kuckartz, 2016; Mayring, 2001, 2015).



**Tab. 7:** Zuordnung der deduktiv zu entwickelnden Kategorien zur Erfassung des Konstrukts Analysekompetenz

Konstrukt		Analysekompetenz	
Dimension	inhaltliche Dimension		formale Dimension
Faktoren	didaktisch-methodisches Wissen	fachliches (incl. fachdidaktisches) Wissen	5 Stufen der Komplexität der Informationsverarbeitung
Kategorien (beispielhafte Nennungen)	Zielorientierung	Historische Genese von Fachwissen	Verstehenstiefe – Level 1
	Exemplarizität	Experimentieren als zentrale Fachmethode	Verstehenstiefe – Level 2a
	...	...	...

### 2.5.2 Bestimmung der Analyseeinheiten

Nachdem Kategorien definiert, beschrieben und mit Ankerbeispielen konkretisiert sind, müssen die *Analyseeinheiten* bestimmt werden.<sup>52</sup> Diese geben an, *welche* Textteile mit Hilfe *welcher* Kategorien analysiert werden sollen (s. Mayring, 2012, S. 29). Hinsichtlich dieses Selektionsprozesses sind mindestens drei Analyseeinheiten zu unterscheiden: Auswertungseinheit, Kodiereinheit, Kontexteinheit.

- *Auswertungseinheit*: Vor Beginn der Analyse des Textmaterials ist die grundsätzliche Entscheidung zu treffen, ob das *gesamte* vorliegende Textmaterial oder nur *bestimmte* (als besonders relevant erachtete) Ausschnitte mit dem zur Anwendung kommenden „Kategoriensystem in Verbindung gebracht werden“ (ebd.) sollen. Dazu wird die Auswertungseinheit bestimmt, die festlegt, wie umfangreich die zu analysierenden „Materialportionen“ (ebd.) sein sollen. Im vorliegenden Fall orientieren sich die Interviewführung und der Interviewleitfaden am Ablauf der videografierten Unterrichtsstunde. Um nun sehen zu können, inwiefern die ProbandInnen den *gesamten* Unterrichtsprozess im Blick haben, wurde festgelegt, *jedes* Interview *komplett* heranzuziehen.<sup>53</sup>
- *Kodiereinheiten*: Sie stellen die *minimalen* Textbestandteile dar, die einer Kategorie zugeordnet werden sollen. Nach Mayring bestimmen sie „die Sensibilität des Verfahrens“ (Mayring, 2012, S. 29), da es sich um das kleinste bedeutungstragende Merkmal handelt, aufgrund dessen die Zuordnung einer Kategorie zu einer Textstelle erfolgt

<sup>52</sup> Der Begriff *Analyseeinheit* wird in der Methodenliteratur nicht einheitlich verwendet. Während Mayring den Begriff als Oberbegriff für die Auswertungs-, Kontext- und Kodiereinheit gebraucht (s. Mayring, 2015), greift Kuckartz auf eine eigene Systematik zurück. Bei ihm werden Analyseeinheiten als Teile einer *Auswahleinheit* aufgefasst. Dabei stellen für Kuckartz Auswahleinheiten „die Grundeinheit einer Inhaltsanalyse dar“, die „nach einem bestimmten Auswahlverfahren ... aus der Grundgesamtheit (d.h. der Menge aller potenziellen Untersuchungsobjekte) für die Inhaltsanalyse ausgewählt“ (Kuckartz, 2016, S. 30) werden.

<sup>53</sup> Darüber hinaus wird mit der Auswertungseinheit festgelegt, „welche Textteile jeweils nacheinander ausgewertet werden“ (Mayring, 2015, S. 61). Mithilfe der „Grundeinheit einer Inhaltsanalyse“ (Kuckartz, 2016, S. 30) wird entschieden, welche Textbestandteile in welcher Reihenfolge zur Analyse herangezogen werden. Die Entscheidung, in welcher Reihenfolge das Material analysiert wird, kann auf unterschiedlichen Wegen realisiert werden. Beispielsweise könnten die ExpertInnen- und Studierendeninterviews abwechselnd analysiert werden. Ich habe festgelegt, dass zunächst die ExpertInneninterviews und anschließend erst die Studierendeninterviews ausgewertet werden sollen, da die ExpertInneninterviews (vermutlich) mehr inhaltliche Facetten aufweisen, sodass die deduktiv aufgestellten Kategorien eine bessere Testung erfahren und abschließend mit den Studierendeninterviews abgeglichen werden können.

- (s. Kuckartz, 2016). Eine Kodiereinheit muss eine hinreichend klare Aussage über eine wahrgenommene Situation/Handlung beinhalten, die unter dem Gesichtspunkt der Lernwirksamkeit beurteilt wird.
- *Kontexteinheiten*: Kategorien können so komplex sein, dass sie nicht durch einzelne Kodiereinheiten abgedeckt werden können. Dann ist es erforderlich, weitere Textstellen heranzuziehen. Diese zusätzlich herangezogenen Textstellen bilden zusammen eine Kontexteinheit (s. Mayring, 2012, S. 29), die Kuckartz als größte Einheit versteht, „die hinzugezogen werden darf, um eine Codiereinheit zu verstehen und richtig zu kategorisieren“ (Kuckartz, 2016, S. 44). Die Festlegung einer Kontexteinheit sei an einem Beispiel verdeutlicht: Bei der Analyse von Unterricht spielt der Komplexitätsgrad zu verarbeitender Informationen, der hier als *Verstehenstiefe* (VT) operationalisiert wird (s. Kap. 3.3), eine zentrale Rolle. Zur Erfassung unterschiedlicher Grade dieser Verstehenstiefe habe ich Kategorien mit vier verschiedenen Levels definiert. Die Kategorie mit dem höchsten Komplexitätslevel heißt „Verstehenstiefe – Level 3a“ (VT-3a) und ist durch drei Kriterien charakterisiert: (1) Der Proband bzw. die Probandin muss das *Ziel* der Unterrichtsstunde umreißen bzw. benennen. (2) Die Beschreibung der Zielstellung muss auf das (vermutete) *Vorwissen* der SchülerInnen bezogen werden. (3) Die *Verbindung* von Zielstellung und Annahmen über das Vorwissen ermöglicht eine (vorerst hypothetische) Einschätzung der Realisierbarkeit der Zielstellung. Damit solche komplexen Kategorien wie die hier am Beispiel erläuterte Kategorie VT-3a auch zur Anwendung kommen können, habe ich mich entschieden, dass eine Kontexteinheit entweder aus einem (längeren) zusammenhängenden Sinnabschnitt bestehen kann oder aber aus mehreren (nicht unmittelbar aufeinanderfolgenden) Textbestandteilen, durch die alle drei Kriterien über mehrere Textstellen hinweg abgedeckt werden.

### 2.5.3 Konsensuelles Kodieren zur Überprüfung der Bildung und Anwendung von Kategorien

Die Kodierung von Textstellen ist ein *interpretativer* Akt. Wie bei jeder anderen Interpretation stellt sich daher auch bei der Kodierung im Rahmen einer Qualitativen Inhaltsanalyse die Frage, ob die jeweilige Interpretation dem Material gerecht wird. Denn Faktoren wie etwa die spezifischen inhaltlichen Interessen der kodierenden Person, seine theoretische Sicht der Dinge, seine Vorlieben und Vorurteile oder das eher unbewusste Einfließen von nicht explizit geklärten Prämissen können zu entsprechenden subjektiven Verzerrungen führen.

Deshalb ist es notwendig, den Prozess der Kodierung am Maßstab *intersubjektiver Überprüfbarkeit* auszurichten (s. z.B. Mayring, 2015; Steinke, 1999, 2009). In der Regel wird diese Forderung nach Intersubjektivität durch ein entsprechendes *Konsensverfahren* einzulösen versucht (s. z.B. Flick, 2009; Gläser-Zikuda, 2015; Kuckartz, 2016; Mayring, 2002, 2010, 2015).

Im Falle meiner Qualitativen Inhaltsanalyse waren an diesem Konsensverfahren vier Personen beteiligt (drei studentische Hilfskräfte und ich).<sup>54</sup> Voraussetzung für die Mitwirkung an der Konsensfindung ist eine hinreichende „Interpretationskompetenz“ (Kuckartz, 2016, S. 44), über die Studierende in der Regel zunächst nicht verfügen. Deshalb muss diese Kompetenz schrittweise über ein entsprechendes *Trainingsprogramm* (s. Anhang E, S. 219) aufgebaut werden (s. Kuckartz, 2016).

Zunächst wurden die Studierenden in die Fragestellung meiner Arbeit (Validierung des Konstrukts Analysekompetenz) eingeführt. Anschließend erhielten sie die Gelegenheit, sich das Video mehrfach anzusehen, um sich mit den inhaltlichen Details der Unterrichtsstunde vertraut zu machen. Dann begann die eigentliche Einführung in das Kategoriensystem und die ihm zugrunde liegenden Dimensionen und Faktoren. Anhand der schriftlichen Dokumentation des bestehenden Kategoriensystems wurden Schritt für Schritt die einzelnen Kategorien mit ihren Definitionen, Beschreibungen und Ankerbeispielen durchgearbeitet und gemeinsam diskutiert, um Fragen und eventuelle Verständnisschwierigkeiten zu klären.

Nach dieser theoretischen Einführung wurden die Studierenden im praktischen Gebrauch der Kategorien trainiert. Dazu wurden zwei Transkripte ausgewählt, an denen zunächst gemeinsam und danach selbstständig die Text-Kategorie-Zuordnung geübt werden konnte. Durch diese Übungen gewannen die Studierenden Schritt für Schritt eine hinreichende Sicherheit im Umgang mit *allen* Kategorien. Sobald genügend Sicherheit im Umgang mit den Kategorien bestand, konnte die Übung auf weitere Transkripte ausgedehnt und schließlich die Trainingseinheit abgeschlossen werden.

Nach Abschluss des Trainings begann der eigentliche Prozess der *konsensuellen Kodierung*. Dabei machte es die Komplexität des Konstrukts Analysekompetenz erforderlich, dass das konsensuelle Kodieren nicht mit allen Kategorien zugleich erfolgen konnte, sondern die *einzelnen* Dimensionen mit den dazugehörigen Faktoren in drei aufeinander folgenden Phasen bearbeitet wurden: (1) Zunächst wurden die Transkripte auf Kategorienvorfälle hinsichtlich der inhaltlichen Dimension (didaktisch-methodisches Wissen) getestet. (2) Im Anschluss daran kamen die Kategorien für das fachliche (bzw. fachdidaktische) Wissen zum Einsatz. (3) Zum Abschluss wurden die Transkripte bezüglich der formalen Dimension (Verstehenstiefe) untersucht.

Vom Ablauf her waren die einzelnen Phasen identisch: Zu Beginn der konsensuellen Kodierung für die entsprechende Dimension (bzw. des dazugehörigen Faktors) erfolgte ein erster Materialdurchlauf (*Pilottestung*) an zufällig ausgewählten Transkripten. Dieser Schritt diente einerseits dazu, sich mit den Kategorien (incl. ihrer Definitionen, Beschreibungen und Ankerbeispielen) vertraut zu machen und andererseits die Interaktion und das Auswertungsverfahren in der ForscherInnengruppe aufeinander abzustimmen, um die Prüfung der Kategorientauglichkeit vornehmen zu können. Dazu wurden mehrere Sitzungen (*Kodierkonferenzen*) einberufen.

Bei dieser Prüfung kann es vorkommen, dass sich bestimmte (bisher geltende) *Kodierregeln* als nicht tragfähig erweisen. Das ist auch bei uns der Fall gewesen, sodass Modifikationen der Kodierregeln notwendig wurden (s. Schmidt, 1997, S. 556ff.). Die auf diese

54 Diese Form des Konsensverfahrens wird abgegrenzt von anderen Konsensverfahren wie der *kommunikativen* Validierung oder der *argumentativen* Validierung. Während bei ersterer ein „Konsens zwischen Forschenden und Beforschten“ (Gläser-Zikuda, 2015, S. 127) hergestellt wird, wird bei der argumentativen Validierung ein „Konsens zwischen Forschenden und außenstehenden Personen“ (ebd.) angestrebt.

Weise präzisierten Interpretationsentscheidungen für die jeweiligen Kategorien wurden anschließend in einem *Kodierleitfaden* (s. die Kategoriendarstellungen in Kapitel 3) festgehalten.<sup>55</sup> Anhand dieses Kodierleitfadens erfolgte dann wiederum ein erneuter und gleichzeitig abschließender *Materialdurchgang*<sup>56</sup>, mit dem die endgültigen Kodierungen vorgenommen wurden.

Der Abschluss des konsensuellen Verfahrens war dann erreicht, wenn die Beteiligten der Ansicht waren, dass *sämtliche* Fundstellen im Text markiert und mit den entsprechenden Kategorien versehen wurden. Am Ende dieses langen Prozesses entstand somit eine konsensuell gefundene *Musterlösung* (s. Lissmann, 2008).

## 2.5.4 Prüfung der Reliabilität

Die wesentliche Funktion des im vorangegangenen Abschnitt (2.5.3) beschriebenen Konsensverfahrens besteht darin zu prüfen, ob die gebildeten Kategorien hinreichend präzise und trennscharf sind und ob sie sich zur Kodierung spezifischer Textstellen eignen. Fällt diese konsensuelle Prüfung positiv aus, ist zugleich ein gewisses Maß an *Intersubjektivität* für den Prozess der Kodierung gegeben.

Trotz eines solchen Konsenses ist aber nicht auszuschließen, dass die Kodierungsergebnisse durch gewisse *Verzerrungen* zustande gekommen sein können. So könnte z.B. durch das Kompetenzgefälle zwischen wissenschaftlichem Mitarbeiter und studentischen Hilfskräften ein Gruppendruck entstanden sein, durch den Argumente des Mitarbeiters allzu schnell akzeptiert worden sind oder durch den mögliche Einwände der Studierenden, die sich bei näherer Betrachtung als berechtigt erwiesen hätten, zu zaghaft formuliert oder erst gar nicht ausgesprochen wurden.

Aufgrund solcher potenziellen Einschränkungen der Intersubjektivität erscheint es wichtig, den erzielten Konsens selbst noch einmal in Frage zu stellen. Das geschieht dadurch, dass *außenstehende* Personen, d.h. Personen, die nicht am Konsensverfahren beteiligt waren, gebeten werden, das Material ebenfalls zu kodieren. Im Idealfall müssten sie spezifische Textstellen auf die gleiche Weise kodieren, wie diejenigen KodiererInnen, die bereits am vorangegangenen Konsensverfahren beteiligt waren. Dieser Idealfall würde dann für die *Reproduzierbarkeit* der Kodierungen durch Andere sprechen und einen weiteren Beleg für die *Intersubjektivität* des angewandten methodischen Verfahrens darstellen (s. Mayring, 2010, 2015).

Für dieses Kriterium der Reproduzierbarkeit wird in der quantitativen Forschungsmethodik üblicherweise der Begriff der *Reliabilität* benutzt, der Ausdruck für das Maß ist, „wie genau der Test das misst, was er misst“ (Rost, 2004, S. 33). Die Güte der Reliabilität wird numerisch durch einen entsprechenden Koeffizienten angegeben.<sup>57</sup> Da bei meinem qualitativen Vorgehen aber weder numerische Werte noch Rangordnungen vorliegen, ist es

55 Der Kodierleitfaden ist das „zentrale Hilfsmittel“ bei der *strukturierenden* Inhaltsanalyse, da er „für jede Kategorie eine Definition, typische Textpassagen als Ankerbeispiele und Kodierregeln zur Abgrenzung zwischen den Kategorien enthält“ (Mayring & Fenzl, 2014, S. 548).

56 Diese eingebauten „Rückkopplungsschleifen“ und die dazugehörigen „Pilottestungen“ sollen gewährleisten, dass jede Kodierentscheidung auf eine getestete Regel zurückzuführen ist und eine Überarbeitung der Kodierregeln sofort auf den Kodierprozess übertragen wird, sodass erst nach den Pilotdurchgängen „von einer bewährten, verlässlichen Methode ausgegangen werden kann“ (Mayring, 2010, S. 603).

57 Solch eine Maßzahl stellt beispielsweise der *Kuder-Richardson-Koeffizient* dar, mit dem festgestellt werden kann, welche Items bzw. Aufgaben zur Messung einer einzelnen latenten, d.h. nicht direkt beobachtbaren,

notwendig, andere geeignete Kriterien für die Verlässlichkeit der „Messungen“ zu wählen. Diese Kriterien werden in der Forschungsliteratur als *Intra-* und *Intercoderreliabilität* (s. z.B. Gläser-Zikuda, 2015; Lauf, 2001; Mayring, 2010, 2014, 2015) bezeichnet. Da im Falle meiner Qualitativen Inhaltsanalyse *zwei* außenstehende Kodiererinnen<sup>58</sup> die Reliabilitätsprüfung vornehmen sollten (es könnten auch mehrere sein), bedeutet dies für die Intra- und Intercoderreliabilität:

- 1) *Intracoderreliabilität*: Jeder der beiden Kodiererinnen muss in der Lage sein, Textstellen zutreffend kodieren zu können. Im Idealfall bedeutet dies, dass *jede* Kodiererin ein und derselben Textstelle zu *verschiedenen* Zeitpunkten die gleiche Kodierung zuweist (s. Lauf, 2001; Mayring, 2010). Die Absicherung dieser *individuellen* Stabilität in der Kodierung stellt eine entscheidende *Voraussetzung* für die darüber hinausgehende Intercoderreliabilität<sup>59</sup> dar.
- 2) Zur Ermittlung der *Intercoderreliabilität* werden die Kodierungsentscheidungen der beiden Kodiererinnen *miteinander* verglichen und überprüft, ob sie „nachweislich zu ähnlichen Ergebnissen“ (Mayring, 2009, S. 471) gekommen sind. Der Idealfall läge hier also vor, wenn *beide* Kodiererinnen ein und derselben Textstelle ein und dieselbe Kategorie zuordnen.<sup>60</sup>

Sowohl die Intra- als auch die Intercoderreliabilität sind allerdings an eine hinreichende *Interpretationskompetenz* der Kodiererinnen gebunden, die über entsprechende Übungen gesichert werden muss. Im Prinzip ist hierzu das gleiche Trainingsverfahren anzuwenden, mit dem bereits die am Konsensverfahren beteiligten Kodierer geschult wurden. Das heißt im Einzelnen: theoretische Einführung in die Fragestellung meiner Arbeit und in die Konzeption des theoretischen Konstrukts Analysekompetenz; Vorstellung des Kategoriensystems; zunächst angeleitetes und danach selbstständiges Training der Kategorienanwendung (s. Trainingsprogramm im Anhang E, S. 219). Erst wenn durch ein solches Trainingsverfahren sichergestellt ist, dass die Kodiererinnen hinreichend gute Werte erzielen, kann mit der eigentlichen Kodierung begonnen und später die Intra- und Intercoderreliabilität ermittelt werden (s. Früh, 2015).<sup>61</sup>

Beim vorausgegangenen konsensuellen Verfahren ging es darum, (möglichst) alle relevanten Textstellen mit Kategorien zu kodieren. Insgesamt konnten auf diese Weise 708 Kodierungen (Kategorienvorfälle) vorgenommen werden. Aus zeitökonomischen

---

Variable dienen. Somit kann mithilfe des Koeffizienten angegeben werden, mit welchen Items die gleichen bzw. ähnlichen Variablen gemessen werden (s. Bortz & Döring, 2006).

58 Bei den Kodiererinnen handelt es sich um zwei weibliche Personen, sodass ich die weibliche Sprachform verwende.

59 Mit der Intercoderreliabilität wird im Grunde genommen die *Auswertungsobjektivität* gemessen und überprüft, inwieweit „die Unabhängigkeit der Ergebnisse von der untersuchenden Person“ (Mayring, 2015, S. 124) gewährleistet ist, daher gilt sie auch eher als Reliabilität im weiten Sinne (s. Bortz & Döring, 2006; Mayring, 2015).

60 Zur Berechnung der Inter- und Intracoderreliabilität werden im Normalfall nur die Ergebnisse der außenstehenden Kodiererinnen miteinander verglichen. Erst im Rahmen der Überprüfung der *Forscher-Kodierer-Reliabilität* (s. unten) werden sowohl das Urteil der Kodiererinnen als auch das des Forschers zueinander in Beziehung gesetzt (s. Früh, 2015).

61 Die Kodierungen wurden in separaten Räumen durchgeführt. Als einziges Hilfsmittel diente dabei der Kodierleitfaden, sodass der Kodierungsprozess unter identischen Bedingungen erfolgte und damit die Durchführungsobjektivität gewährleistet wurde (s. Diekmann, 2009; Früh, 2015; Mayring, 2015).

Gründen kann diese große Zahl an Fundstellen nicht für die darauf folgende Reliabilitätsprüfung in Betracht gezogen werden. Stattdessen sollten daraus etwa zehn Prozent per Zufall selektiert und überprüft werden.

Diese Zufallsziehung soll letztlich dazu führen, dass alle Kategorien des Kategoriensystems gleichmäßig zum Zuge kommen. Denn dann müssen die beiden Kodiererinnen zeigen, dass sie nicht nur einige Kategorien, sondern *alle* richtig verstanden haben und anwenden können. Praktisch sieht das Auswahlverfahren so aus: Insgesamt sind 708 Zettel (Lose) vorhanden, die von 1 bis 708 durchnummeriert sind. Jede Zahl steht für eine der 708 Textstellen, denen im Konsensverfahren eine Kategorie zugeordnet wurde (359 Fundstellen für das didaktisch-methodische Wissen; 179 Fundstellen für das fachliche [incl. fachdidaktische] Wissen; 170 Fundstellen für die Stufen der Komplexität der Informationsverarbeitung [Verstehenstiefe]). Die Zettel werden nun auf drei Lostöpfe verteilt, wobei jeder Lostopf einen der drei Faktoren der beiden Dimensionen von Analysekompetenz vertritt (s. die folgende Tab. 8).

**Tab. 8:** Übersicht zu Fundstellen der jeweiligen Faktoren

Dimension	Faktoren	Kategorien-anzahl	Fundstellen	Lostopf
Inhaltlich	Didaktisch-methodisches Wissen	12	359	1
	Fachliches (incl. fachdidaktisches) Wissen	7	179	2
Formal	Stufen der Komplexität der Informationsverarbeitung	5	170	3
Gesamt		24	708	

In einem Lostopf befinden sich also alle Kategorienvorfälle für das didaktisch-methodische Wissen (359 Fundstellen), in einem zweiten für das fachliche (incl. fachdidaktische) Wissen (179 Fundstellen) und im dritten für die Stufen der Komplexität der Informationsverarbeitung (170 Fundstellen). Auf diese Weise werden alle 24 Kategorien des Kategoriensystems und alle identifizierten Fundstellen der Überprüfung (Ziehung) zugänglich gemacht.

Bei der Ziehung selbst wird so vorgegangen, dass alle 24 Kategorien mindestens *dreimal* in der endgültigen Auswahl des Textmaterials vertreten sein sollen, sodass 72 Fundstellen (ca. 10% aller Kategorienvorfälle) ausgewählt werden müssen.

Der sich anschließende *Prozess* der Kategorisierung durch die beiden Kodiererinnen sieht so aus, dass ihnen die per *Zufallsverfahren* selektierten Textstellen vorgelegt werden. Sie sind durch eine am Rand markierte Linie bestimmter Länge vorgegeben, so dass den Kodiererinnen angezeigt wird, welche Textstelle gemeint ist. Diesen markierten Textstellen sollen die Kodiererinnen dann die zutreffenden Kategorien zuordnen, indem sie das dafür vorgesehene Kürzel neben der Linie vermerken.

Die „Quote der übereinstimmenden Einschätzungen“ (Mayring, 2015, S. 127) zwischen den Kodiererinnen lässt sich dann nach der sogenannten *Holsti-Formel* berechnen. Dazu werden die Übereinstimmungen addiert und mit dem Faktor *n* (Anzahl der Kodierer) multipliziert, um den so ermittelten Wert dann durch die Summe aller Kodierungen (die von



Kodierer 1 und 2 vorgenommen wurden) zu dividieren. Multipliziert man diesen Wert mit dem Faktor 100, erhält man die *prozentuale Übereinstimmung* zwischen beiden Kodierenden.

Dieser Koeffizient erweist sich aber insofern als fehleranfällig, weil die Übereinstimmungen durch Raten und somit *zufällig* zustande gekommen sein könnten (s. Mayring, 2015, S. 128; Zierer et. al., 2013).<sup>62</sup> Um dieses Maß an Zufälligkeiten einzuschätzen, wird in der Regel ein „zufallskritischer Koeffizient“ (Lissmann, 2008, S. 173) herangezogen, wie ihn beispielsweise *Cohens Kappa* darstellt. Der Grundgedanke hierbei ist es, die Anzahl der Ausprägungen um die zufälligen Übereinstimmungen zu korrigieren (s. z.B. Bortz & Schuster, 2010, S. 505; Lissmann, 2008, S. 173f.; Mayring, 2015, S. 128; Wirtz & Caspar, 2002). Zur Berechnung von Cohens Kappa greife ich den Vorschlag von Brennan und Prediger auf, mit dem die Anzahl der vorhandenen Kategorien berücksichtigt wird (s. Brennan & Prediger, 1981).<sup>63</sup> Das Ergebnis dieser Rechnung liegt im Wertebereich zwischen 0 und 1, wobei ein Kappa-Koeffizient zwischen .60 und .75 als ausreichend gilt (s. z.B. Cohen, 1988, 1992; Fleiss & Cohen, 1973; Mayring, 2000, 2009).

### 2.5.5 Prüfung der Validität

Mit der in Kapitel 2.5.4 beschriebenen Bestimmung der Reliabilität lassen sich vor allem Aussagen über die Zuverlässigkeit und Genauigkeit der „Messungen“ machen (s. Atteslander, 2010, S. 209; Schermelleh-Engel & Werner, 2012, S. 120). Die ermittelte Reliabilität ist aber nur eine *notwendige* Bedingung für die Güte eines Testverfahrens (s. Lauf, 2001). Als *hinreichende* Bedingung muss zusätzlich ein Erweis für dessen *Validität* erbracht werden (s. z.B. Früh, 2015). Diese „Validitätsprüfung“ (Früh, 2015, S. 189) entspricht dem *traditionellen* Verständnis von Validität und wird über die Frage definiert, „inwieweit der Test das misst, was er messen soll“ (Rost, 2004, S. 33). Grundgedanke dabei ist, dass „die vom Forscher letztlich als valide erachteten Operationalisierungen eines als valide begründeten Konstrukts ... von den Codierern auch adäquat umgesetzt werden“ (Früh, 2015, S. 190) müssen. Denn nur wenn der Forscher „seine *als valide gesetzte Interpretationsweise* angemessen den Codierern vermitteln kann, sind auch die produzierten Daten in diesem Sinne valide“ (ebd., kursiv im Original).

Im Rahmen meiner Qualitativen Inhaltsanalyse wird die Validität nach Früh (2015, S. 189ff.) in der Weise geprüft, dass die von den externen Kodierenden vorgenommenen Kategorisierungen mit den im vorausgegangenen Konsensverfahren ermittelten Kategorisierungen verglichen werden. Sind die beiden Kategorisierungen deckungsgleich, so ist anzunehmen, dass die externen Kodierenden und die am Konsensverfahren beteiligten Personen ein *gemeinsames* Verständnis des Kategoriensystems, d.h. der Definitionen und Beschreibungen der Kategorien samt zugehöriger Ankerbeispiele, teilen. Insgesamt ist

62 Früh sieht dies aus drei Gründen nicht als schwerwiegend an. *Erstens* sei ein Zufallstreffer auch eine richtige Kodierung; *zweitens* könnten die Zufälle desto mehr minimiert werden, je größer die Textmenge sei und *drittens* sollten Koeffizienten mit einem niedrigen Wert ( $CR = .50$ ) sowieso abgelehnt werden (s. Früh, 2015, S. 188).

63 Die Berechnung sieht wie folgt aus:  $Kappa = [gemessene\ Übereinstimmung - 1/k] / [1 - 1/k]$ . Das Kürzel  $k$  steht dabei für die Anzahl der Kategorien. Für den Faktor des didaktisch-methodischen Wissens müsste daher bei der Berechnung des Kappa-Wertes ein  $k = 12$  eingesetzt werden, da der Faktor des didaktisch-methodischen Wissens durch 12 Kategorien repräsentiert wird (s. Kap. 3.1).

dann auch davon auszugehen, dass die Personen in beiden Gruppen eine gemeinsame theoretische Basis haben. In meinem konkreten Fall wäre dies das Verständnis des zweidimensionalen Konstrukts Analysekompetenz mit den zugehörigen drei Faktoren. Als Maß für die Validität, also für die Übereinstimmung zwischen den Kategorisierungen der externen Kodiererinnen mit denen der im Konsensverfahren beteiligten Personen, setze ich ebenfalls als Richtwert einen Kappa-Koeffizienten zwischen .60 und .75 an. Ist dieses Minimum erreicht bzw. überschritten, gehe ich davon aus, dass das „Testverfahren“ das misst, was es messen soll.<sup>64</sup>

## 2.6 Ablaufmodell der Qualitativen Inhaltsanalyse

In den vorangegangenen Teilkapiteln wurde ausführlich beschrieben, wie das Vorgehen meiner *inhaltlich-strukturierenden* Qualitativen Inhaltsanalyse realisiert wird. An dieser Stelle möchte ich die vielfältigen inhaltlichen und methodischen Überlegungen bündeln und die wesentlichen Aspekte in einer Tabelle zusammenführen. Diese Tabelle stellt gleichzeitig das *Ablaufmodell* meiner Qualitativen Inhaltsanalyse dar, anhand dessen das *systematische, regelgeleitete* und an den Maßstäben der *intersubjektiven Nachvollziehbarkeit* orientierte Vorgehen erkennbar ist (s. z.B. Mayring, 2012, 2015; Steigleder, 2008; Steinke, 1999, 2009).<sup>65</sup>

In Anlehnung an Mayring (2015) und Schreier (2014) sind in der linken Spalte der Tabelle 9 die „allgemeinen“ Ablaufschritte einer *inhaltlich-strukturierenden* Qualitativen Inhaltsanalyse angeführt. In der rechten Spalte werden diese Schritte für meine Arbeit konkretisiert.

64 Zur Überprüfung des gemeinsamen Verständnisses werden die Kodierungen der beiden externen Kodiererinnen mit der bereits im Konsensverfahren ermittelten „Standardkodierung“ (Lissmann, 2008, S. 171) verglichen. Dazu wird eine „Kodierkonferenz“ (Mayring, 2010, S. 604) einberufen, zu der die beiden Kodiererinnen und der Forscher anwesend sind, um etwaige Zweifelsfälle (insb. Nicht-Übereinstimmungen) zu besprechen. Eine Nicht-Übereinstimmung liegt aber erst dann vor, wenn die beiden außenstehenden Kodiererinnen dem Forscher entweder Fehlkodierungen oder Abweichungen vom Kodiermanual nachweisen können (s. Mayring, 2010, 2015).

65 Dieses Ablaufmodell ist an ein konkretes Verfahren der Qualitativen Inhaltsanalyse gebunden und die bisher verwendete Bezeichnung „Qualitative Inhaltsanalyse“ verweist auf allgemeiner Ebene „nur“ auf die *Methode*, die eingesetzt wird, um Datenmaterial auszuwerten. Diese Methode lässt sich aber in mehrere Grundformen bzw. Varianten unterteilen, die als spezifisch qualitativ-inhaltsanalytische *Verfahren* in einer Untersuchung zum Einsatz gebracht werden (s. Stamann et al., 2016). Es lassen sich eine Vielzahl von spezifischen qualitativ-inhaltsanalytischen Verfahren unterscheiden. Beispielsweise unterscheidet Mayring prinzipiell drei Grundformen voneinander (s. Mayring, 2010, S. 602; 2012): die *zusammenfassende*, die *explikative* und die *strukturierende* Inhaltsanalyse. Dabei wird die strukturierende Inhaltsanalyse wiederum in vier Unterformen unterteilt: die *formale*, die *inhaltliche*, die *typisierende* und die *skalierende* Strukturierung. Im Gegensatz dazu unterscheidet Kuckartz (2016) auch drei Varianten qualitativ-inhaltsanalytischen Vorgehens: die *inhaltlich-strukturierende*, die *evaluative* und die *typenbildende* Qualitative Inhaltsanalyse.

Welches qualitativ-inhaltsanalytische Verfahren verwendet wird, ist von der Fragestellung und dem Erkenntnisinteresse der Untersuchung abhängig und wird bspw. durch die Art der verwendeten Kategorien (formal, inhaltlich, skalierend) oder den Modus der Kategorienbildung (deduktiv, induktiv, Mischform) bestimmt (s. z.B. Stamann et al., 2016).



**Tab. 9:** Ablaufmodell der Qualitativen Inhaltsanalyse

Schritte der Inhaltsanalyse	Konkrete Umsetzung in der Untersuchung
Präzisierung und theoretische Begründung der Fragestellung	<p>Die zentrale Fragestellung der Untersuchung liegt darin, ob sich die manifesten Daten meiner qualitativen Studie als valide Belege für die faktorielle Struktur des Konstrukts Analysekompetenz interpretieren lassen.</p> <p>Die theoretischen Annahmen, die dem Konstrukt Analysekompetenz von Plöger et al. (2015) zugrunde liegen, werden expliziert.</p> <p>Dabei wird das Konstrukt Analysekompetenz als zweidimensionales Konstrukt interpretiert, das sich aus einer inhaltlichen und einer formalen Dimension zusammensetzt. Die inhaltliche Dimension wird in zwei Faktoren unterteilt: didaktisch-methodisches Wissen einerseits und fachliches (incl. fachdidaktisches) Wissen andererseits. Die formale Dimension ist durch drei Faktoren gekennzeichnet: Stufe 1 und 2 des fünfstufigen Kompetenzmodells von Plöger und Scholl repräsentieren die Analytische Kompetenz, die Stufen 3 und 4 umfassen die Synthetische Kompetenz und die Stufe 5 steht für den Faktor der Prozesskompetenz.</p> <p>Des Weiteren basierte die Untersuchung von Plöger und Scholl (2014a) auf dem sogenannten <i>well-known-group-Vergleich</i> (s. z.B. Cronbach &amp; Meehl, 1955, DeVellis, 1991; Hattie &amp; Cooksey, 1984; Rubin &amp; Babbie, 2015). Dabei geht es um den Vergleich von zwei Gruppen, von denen angenommen wird, dass sie sich hinsichtlich eines infrage stehenden Merkmals (hier Analysekompetenz) voneinander unterscheiden.</p> <p>Plöger und Scholl (2014a) konnten zeigen, dass Fach- und SeminarleiterInnen (ExpertInnen) die höchsten Stufen der Analysekompetenz erreichten, dagegen die Studierenden (NovizInnen) nur die niedrigeren Stufen. Zwischen diesen beiden Gruppen lagen die ReferendarInnen und praktizierenden Lehrpersonen. Dieses Ergebnis wird auch in meiner Untersuchung bei der Zusammensetzung der Stichprobe (ExpertInnen, NovizInnen) berücksichtigt, um Unterschiede in den Analyseleistungen sichtbar machen zu können.</p>
Auswahl und Charakterisierung des Materials	<p>Die zu analysierenden Daten werden mittels fokussierter, leitfadengestützter Interviews erhoben.</p> <p>Die Stichprobe umfasst 18 Personen, die sich hinsichtlich des Grades ihrer Analysekompetenz (ExpertInnen/NovizInnen) und ihrer Fachzugehörigkeit unterscheiden (Fach Physik; ein anderes naturwissenschaftliches Fach und/oder Mathematik; keines dieser Fächer).</p> <p>Die Stichprobenziehung erfolgte nach diesen beiden Kriterien und kann somit als bewusst bzw. absichtsvoll bezeichnet werden.</p> <p>Der technische Ablauf der Interviews folgt einem viergliedrigen Aufbau (Einstiegsphase; Aufwärmphase; Hauptphase; Ausstiegsphase).</p> <p>Die aufgezeichneten Interviews werden verschriftlicht (transkribiert).</p>
Bestimmung der Richtung der Analyse	Zur Überprüfung der faktoriellen Struktur des Konstrukts Analysekompetenz werden Hypothesen formuliert.
Festlegung der Analyseinheiten	<p>Alle Interviews werden komplett der Analyse unterzogen (Auswertungseinheit). Eine Kategorie wird aber nur dann vergeben, wenn die Aussagen der ProbandInnen eine hinreichende Begründung enthalten (Kodiereinheit). Eine Kontexteinheit kann entweder aus einem (langen) zusammenhängenden Sinnabschnitt oder aber auch aus mehreren kleinen (nicht unmittelbar aufeinanderfolgenden) Textbestandteilen bestehen.</p>

Erstellen des Kategoriensystems (theoriegeleitete Festlegung der Kategorien bzw. Strukturierungsdimensionen)	Die Kategorien werden deduktiv, aus den theoretischen Annahmen heraus, gebildet.
Formulierung des Kodierleitfadens	Die Kategorien werden alle in einem Kodierleitfaden zusammengeführt. Dort finden sich zu jeder Kategorie eine knappe Definition, eine Beschreibung und ein Ankerbeispiel.
Beginn der deduktiven Kategorienanwendung	Die Kodierung des Materials wird mittels Konsensverfahren umgesetzt. An diesem Prozess sind vier Personen beteiligt (drei studentische Hilfskräfte und ich). Die studentischen Hilfskräfte wurden mittels Trainingsverfahren vorab geschult. Nach Abschluss des Trainings beginnt der Prozess der konsensuellen Kodierung. Dabei werden die einzelnen Dimensionen und die dazugehörigen Faktoren nacheinander bearbeitet.
Probekodierungen – Überarbeitung der Kategorien	In mehreren Kodierkonferenzen wird in einem ersten Materialdurchlauf (Pilottestung) die Tauglichkeit der Kategorien überprüft, sodass nach dieser Phase der Kodierleitfaden aktualisiert werden kann.
Endgültiger Materialdurchgang (Hauptkodierung; Fundstellenbezeichnung)	Im Anschluss daran erfolgt ein erneuter, abschließender Materialdurchgang, mit dem die endgültigen Kodierungen vorgenommen werden. Der Abschluss dieses Konsensverfahren ist dann erreicht, wenn alle Beteiligten der Ansicht sind, sämtliche Fundstellen im Text markiert zu haben. Das Ergebnis stellt eine konsensuell gefundene Musterlösung dar.
Überprüfung der Gütekriterien (Reliabilität und Validität)	<p>Zur Prüfung der Reliabilität werden zwei außenstehende Personen, die nicht am Prozess der Kodierung beteiligt waren, gebeten, ca. zehn Prozent des Materials zu kodieren (Intercoderreliabilität).</p> <p>Nach fünf Wochen wird eine Person erneut mit dem Material konfrontiert und muss nochmals die entsprechenden Textstellen kodieren (Intracoderreliabilität).</p> <p>Die Validität (im Sinne Frühs [2015, S. 189ff.], s. S. 69f.) wird überprüft, indem die von den externen Kodiererinnen vorgenommenen Kodierungen mit der im Konsens erstellten „Musterlösung“ abgeglichen werden.</p>
Auswertung und Aufbereitung der Ergebnisse	<p>Nach Überprüfung der Reliabilitäts- und Validitätswerte werden die Kategorienhäufigkeiten (Wissensaktivierungen) dargestellt und im Anschluss entsprechende Textbelege geliefert, die verdeutlichen sollen, welche konkreten Aussagen sich hinter diesen „nackten“ Zahlen verbergen.</p> <p>Anschließend werden die ermittelten Wissensaktivierungen zum einen mit Reliabilitätsanalysen sowie Trennschärfe- und Korrelationsberechnungen überprüft, um beurteilen zu können, ob die ermittelten Kodierungen als ausreichend genaue Messungen interpretiert werden können und zum anderen wird geschaut, ob die Unterschiede zwischen ExpertInnen und NovizInnen signifikant sind.</p>

Nachdem das methodische Vorgehen meiner Untersuchung vorgestellt wurde, sollen im nächsten Kapitel die deduktiv zu bildenden Kategorien hergeleitet und präsentiert werden.

### 3 Theoretische Herleitung der Kategorien

Nachdem im vorangegangenen Kapitel erläutert wurde, wie das methodische Verfahren der Qualitativen Inhaltsanalyse für die Bearbeitung meiner Fragestellung im Allgemeinen verwendet werden soll, geht es in diesem Kapitel darum, das zum Einsatz kommende Kategoriensystem vorzustellen und die einzelnen Kategorien differenziert zu beschreiben.

Die Herleitung der Kategorien erfolgt unter Maßgabe der theoretischen Annahmen, durch die Plöger et al. (2015) Analysekompetenz als zweidimensionales Konstrukt definieren (s. Kap. 1). Dabei repräsentiert die erste Dimension die *inhaltliche* Qualität mit zwei Wissensfacetten als Faktoren, die für die Analyse von Unterricht relevant erscheinen: das *didaktisch-methodische* Wissen einerseits und das *fachliche* (incl. *fachdidaktische*) Wissen andererseits. Die zweite Dimension stellt die *formale* Qualität des Wissens dar, die den Grad der für die Analyseleistungen notwendigen Komplexität der Informationsverarbeitung abbildet und sich in drei Faktoren unterteilen lässt (Analytische Kompetenz, Synthetische Kompetenz und Prozesskompetenz).

Diese zweidimensionale Struktur des Konstrukts Analysekompetenz, die Plöger et al. (2015) via Konfirmatorischer Faktorenanalyse belegen konnten, dient mir nun als *hypothetische* Basis, aus der heraus *deduktiv* Kategorien hergeleitet werden, die das Konstrukt Analysekompetenz in inhaltlicher und formaler Hinsicht repräsentieren bzw. operationalisieren. Insgesamt gesehen sind deshalb für den hier vorliegenden Validierungsversuch und für die dazu zu bildenden Kategorien drei Aspekte von zentraler Bedeutung: (1) das didaktisch-methodische Wissen, (2) das fachliche (incl. fachdidaktische) Wissen sowie (3) die Komplexität der Informationsverarbeitung.

Um die Kategorien für diese drei Aspekte herzuleiten, skizziere ich zunächst den theoretischen Kontext, dem sie entstammen. Anschließend werden die *deduktiv* entwickelten Kategorien definiert, beschrieben und mittels Ankerbeispielen konkretisiert (s. Bortz & Döring, 2006; Flick, 2009; Kuckartz, 2016; Mayring, 2009, 2015).

Die folgende Tabelle 10 gibt vorab einen Überblick über die zu bildenden Kategorien, die in den entsprechenden Teilkapiteln vorgestellt werden:

**Tab. 10:** Kategorien zur Überprüfung des Konstrukts Analysekompetenz

Konstrukt		Analysekompetenz	
Dimension	inhaltliche Dimension		formale Dimension
Faktoren	didaktisch-methodisches Wissen	fachliches (incl. fachdidaktisches) Wissen	5 Stufen der Komplexität der Informationsverarbeitung
Kategorien	Zielorientierung	Physikalische Begriffsbildung	Verstehenstiefe – Level 1
	Strukturierung	Durcharbeiten von Fachbegriffen	Verstehenstiefe – Level 2a
	Klarheit	Experimentieren als zentrale Fachmethode	Verstehenstiefe – Level 2b
	Exemplarizität	(weitere) fachliche Arbeitsweisen	Verstehenstiefe – Level 3a
	Lebensweltbezug des Themas	Historische Genese physikalischen Wissens	Verstehenstiefe – Level 3b
	Verknüpfung mit dem Vorwissen	Produktiver Umgang mit Fehlern	
	Aktivierung/Motivierung zur geistigen Selbsttätigkeit	Lernen als konstruktive Tätigkeit von SchülerInnen	
	Diskursivität des Unterrichtsprozesses		
	Individualisierung/Differenzierung		
	Ergebnissicherung/Erfolgskontrolle		
	Anwendung/Transfer		
	Lernatmosphäre		

### 3.1 Kategorien für didaktisch-methodisches Wissen

#### 3.1.1 Bezüge zum Modell der Struktur- und Prozessqualität von Unterricht

Um nun Kategorien zu entwickeln, mit denen sich das bei der Analyse von Unterricht genutzte *didaktisch-methodische* Wissen von Lehrpersonen erfassen lässt, knüpfe ich zunächst noch einmal an die gegenwärtige Diskussion an (s. Kap. 1.2).

- 1) Im Kompetenzmodell COACTIV werden fünf *Facetten* des pädagogisch-psychologischen Wissens genannt: Wissen über effektive Klassenführung, Wissen über Unterrichtsmethoden, Wissen über Leistungsbeurteilung, Wissen über individuelle Lernprozesse und Wissen über individuelle Besonderheiten von Schülerinnen und Schülern (s. Voss & Kunter, 2011, S. 195ff.).
- 2) In Rahmen von TEDS-M werden vier *Dimensionen* zur Erhebung pädagogischen Wissens unterschieden: Strukturierung von Unterricht, Umgang mit Heterogenität, Klassenführung, Motivation und Leistungsbeurteilung. Diese Dimensionen werden dann in verschiedene *Inhaltsgebiete* aufgeteilt, z.B. wird die Dimension Umgang mit Heterogenität in die beiden Inhaltsgebiete *Differenzierungsmaßnahmen* und *Methodenvielfalt* aufgliedert (s. Blömeke & König, 2010, S. 245).

Beide Konzeptualisierungen zeigen mehr oder weniger deutliche Schnittmengen hinsichtlich der Benennung von Wissensfacetten, die sowohl für eine erfolgreiche Gestaltung wie auch für eine angemessene *Analyse* von Unterricht notwendig erscheinen:

- 1) In beiden Projekten wird bspw. angenommen, dass die Lernprozesse von SchülerInnen durch eine stimmige Strukturierung des Unterrichtsablaufs, durch einen geeigneten Methodeneinsatz, durch das Eingehen auf individuelle Lernschwierigkeiten etc. gefördert bzw. unterstützt werden können.
- 2) Weiterhin sind in beiden Konzeptualisierungen auch begriffliche Übereinstimmungen (z.B. Klassenführung, Methodenvielfalt) bzw. begriffliche Äquivalente (z.B. Wissen über individuelle Lernprozesse einerseits und Heterogenität andererseits) erkennbar.
- 3) Im Hinblick auf *empirische Evidenzen* stützen sich beide (aber auch andere) Konzeptualisierungen auf Metaanalysen, die sich dann zu einem relativ homogenen Bild von „gutem“ Unterricht zusammenschließen lassen (Fraser, Walberg, Welch, & Hattie, 1987; Hattie, 2003, 2013; Helmke, 2002; Scheerens & Bosker, 1997; Seidel & Shavelson, 2007; Slavin, 1994; Wang, Haertel, & Walberg, 1990).<sup>66</sup>

---

66 Trotz dieses relativ konsistenten Bildes über verschiedene Metaanalysen hinweg, besteht aber weiterhin Forschungsbedarf. Lipowsky macht diesen Bedarf am Merkmal der „kognitiven Aktivierung“ fest, das auch ich als Kategorie in meiner Qualitativen Inhaltsanalyse verwenden werde (s. unten 3.1.2, S. 80f.), indem er festhält: „Die Forschungslage zu den Wirkungen eines kognitiv aktivierenden Unterrichts ist noch vergleichsweise dünn“ (Lipowsky, 2006, S. 61). Ein Grund könnte z.B. darin liegen, dass für den empirischen Nachweis der Lernwirksamkeit verschiedener Merkmale ein mehr oder weniger spezifisches experimentelles Design notwendig ist. Das Problem dabei ist, dass entsprechende Ergebnisse nur bedingt generalisiert werden können, je spezifischer solch ein Design angelegt ist. Ein weiterer Grund könnte aber auch in den unterschiedlichen Operationalisierungen der jeweiligen Begriffe liegen. Offensichtlich liegt (immer) noch keine einheitliche Terminologie hinsichtlich der Qualitätsmerkmale von Unterricht vor. Für die Begriffe „kognitive Aktivierung“ (s. z.B. Leuders & Holzäpfel, 2011) bzw. „kognitive Strukturierung“ (s. z.B. Einsiedler & Hardy, 2010), bei denen eigentlich davon ausgegangen werden könnte, dass sie theoretisch geklärt seien, scheint es doch noch erheblichen (theoretischen) Klärungsbedarf zu geben (s. Lipowsky, 2015).

Die hier exemplarisch in Anlehnung an Projekte wie COACTIV und TEDS-M angeführten Facetten didaktisch-methodischen Wissens stellen zunächst einen Minimalkonsens dar, für den eine hinreichende empirische Evidenz vorliegt. Auf diesen Minimalkonsens rekurren auch Plöger und Scholl in dem von ihnen entwickelten Modell der Struktur- und Prozessqualität (s. Abb. 3, S. 24). Im Rahmen ihrer Untersuchung hat sich allerdings gezeigt, dass dieser Minimalkonsens, für eine tiefe Analyse von Unterricht, um weitere Facetten ergänzt werden muss (s. Plöger & Scholl, 2012, 2013, 2014a, 2014b).

Das Struktur- und Prozessmodell diente Plöger und Scholl als theoretische Basis, aus der heraus sie Items für ihren Test entwickelten. Diese Items waren nur lösbar, wenn die ProbandInnen über hinreichend didaktisch-methodisches Wissen verfügen, wie es das Modell repräsentiert. Da ich ebenfalls davon ausgehe, dass die ProbandInnen solche Wissensfacetten für eine angemessene Analyse der ausgewählten Physikstunde (s. Kap. 2.3.2) aktivieren müssen, schließe ich mich diesem Modell für die Durchführung meiner Qualitativen Inhaltsanalyse an und übernehme es als *integrativen* Rahmen, bei dem die „Unterrichtsqualität als wechselseitige Ergänzung von Struktur- und Prozessqualität“ (Plöger & Scholl, 2014b, S. 275) bestimmt wird und die teilweise nebeneinander stehenden Einzelmerkmale von Unterrichtsqualität in einen *systematischen* Zusammenhang gebracht werden.

Die wechselseitige Ergänzung von Struktur- und Prozessqualität besagt, dass die Strukturqualität über die sinnvolle Abstimmung von Zielen/Intentionen/Themen, den Methoden und den Medien unter Berücksichtigung der Lernvoraussetzungen der SchülerInnen gesichert wird. Diese Elemente stellen aber nur den *strukturellen* Rahmen von Unterricht dar. Zur Realisierung von konkreten Lerngelegenheiten werden Prozessmerkmale benötigt, mit denen das Lernen der SchülerInnen optimal gefördert werden soll (s. Kap. 1.2).

Diese Prozessmerkmale wurden bereits in Kapitel 1.2 (s. S. 26) und in Tabelle 10 (s. S. 74) genannt, es wurde jedoch noch nicht gesagt, was die Merkmale im Einzelnen meinen bzw. wie diese bestimmt werden. Diese Bestimmung erfolgt im folgenden Teilkapitel<sup>67</sup>, damit anschließend diese Prozessmerkmale als Kategorien theoretisch abgeleitet und als Analyseinstrumente für meine Qualitative Inhaltsanalyse operationalisiert werden können.<sup>68</sup>

---

67 Auf eine explizite Darstellung der Strukturmomente wird verzichtet, denn diese werden in den Interviews nicht *direkt* besprochen, sondern im Rahmen der Prozessmerkmale *indirekt* behandelt.

68 Die in Tabelle 10 (s. S. 74) angeführten Merkmale, sind diejenigen, die in meiner Untersuchung zum Einsatz kommen. In Abweichung zum Modell der Struktur- und Prozessqualität von Plöger und Scholl habe ich auf einzelne Merkmale (z.B. Aufgabenorientierung) verzichtet, denn nicht alle angeführten Prozessmerkmale (ob positiv oder eher negativ eingelöst) können den realen Verlauf einer Unterrichtsstunde insgesamt bzw. durchgehend bestimmen. Eine Unterrichtsstunde, in der z.B. vorwiegend geübt oder transferiert wird, ist primär nicht durch die Reflexivität von Inhalt und Methode gekennzeichnet. Auch Merkmale wie Klarheit und Strukturierung werden keine beherrschenden Merkmale des Unterrichtsprozesses sein. Dagegen spielen andere Qualitätsmerkmale wie Aufgabenorientierung, Individualisierung, Übung oder Transfer eine entscheidende Rolle. Welche Prozessmerkmale die Lernwirksamkeit einer spezifischen Unterrichtsstunde vorrangig bestimmen, hängt von der spezifischen Faktur der jeweiligen Stunde ab. Daraus resultiert eine wichtige Konsequenz für den folgenden Prozess der Kategorienbildung: Der entscheidende Bezugspunkt liegt in der zu *analysierenden Unterrichtsstunde*; die interviewten ExpertInnen und NovizInnen werden daher nur solche Kategorien zur Anwendung bringen (können), die zu dieser Stunde „passen“. Daher werden nur die Prozessmerkmale definiert und beschrieben, die auf die inhaltliche und methodische Faktur der zu analysierenden Unterrichtsstunde zugeschnitten sind.

### 3.1.2 Prozessmerkmale von Unterricht

#### Zielorientierung

Unterricht ist als intentionales Geschehen immer an *Lernziele* gebunden (Bolhuis, 2003, S. 335; Kyriakides et al., 2013, S. 144; Seidel & Shavelson, 2007, S. 459). Daher sollten die aufgestellten Lernziele hinreichend klar definiert und in Teilziele ausdifferenziert sein. Im tatsächlichen Unterrichtsverlauf stellen Lernziele den „roten Faden“ des Lernprozesses dar. Dabei kann es aber aus vielfältigen Gründen vorkommen, dass sich der Verlauf von Unterricht vom roten Faden dieser Zielsetzungen entfernt. Daher muss die Lehrperson bestrebt sein, den Unterrichtsverlauf stets auf die angestrebten Intentionen hin auszurichten, damit die SchülerInnen diese Intentionen als Ziele ihres *eigenen* Lernprozesses annehmen können (s. z.B. Brophy, 2000; Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002; Seidel, 2014).

Um den Unterrichtsverlauf am roten Faden zu orientieren, stehen der Lehrpersonen verschiedene Möglichkeiten zur Verfügung, die von Herweg (2008, S. 17ff.) als Teilmerkmale (incl. Indikatoren) für Zielorientierung bestimmt werden:

- 1) *Unterrichtsziel* (Indikatoren: Beschreibung der zu behandelnden Inhalte, Ausblick auf den Stunden- und Unterrichtsverlauf, Angabe der zu erreichenden Kompetenzen)
- 2) *explizite Zielorientierung* (Indikatoren: explizite Zielorientierung durch Zielrückgriffe, regelmäßige und abschließende Zusammenfassungen mit Zielklärungen, Zielklärung nach Experimenten)
- 3) *implizite Zielorientierung* (Indikatoren: Sequenzierung und eine kohärente Inhaltserarbeitung, schlüssige Einbettung von Experimenten und anderen Unterrichtselementen, angemessene Informationsfülle, Berücksichtigung des Vorwissens der Lernenden)
- 4) *kontextuelle Anker/Integrationshilfen* (Indikatoren: Integration der neuen Inhalte in bestehende Vorwissensstrukturen, Schaffung von Anwendungsmöglichkeiten – Verknüpfung mit Alltagsphänomenen, Integrationshilfen durch Transparenz des roten Fadens, Einordnung der Inhalte in einen übergeordneten Kontext).

#### Strukturierung

Die zu erlernenden Sinnstrukturen (z.B. mathematische Operationen) stellen für die SchülerInnen relativ komplexe Sachverhalte dar. Deshalb erscheint es zum Aufbau „einer gut organisierten“ (Helmke, 2012, S. 191) Wissensbasis notwendig, den gesamten Lernprozess in inhaltlich aufeinander bezogene *Teilprozesse* zu zerlegen, sodass die angestrebte Komplexität sukzessive aufgebaut werden kann (Leinhardt & Greeno, 1986; Muijs & Reynolds, 2011). Diese Teilprozesse sind wiederum zu einem „Ganzen“ verbunden, insofern jedes Teilziel Bedingungen für das Erreichen weiterer Teilziele bildet (s. z.B. Aebli, 1993a; Brophy, 2000; Drollinger-Vetter, 2011; Good & Brophy, 2000; Helmke & Schrader, 2009; Kleickmann, 2012; König et al., 2014; Kyriakides et al., 2013; Rosenshine & Stevens, 1986; Schröder, 2002; Walberg & Paik, 2000).

Als mögliche Indikatoren für die Strukturierung dienen beispielsweise: Zerlegung des Gesamtthemas in Teilaspekte, stimmige Aufeinanderfolge und Verknüpfung der Unterrichtsphasen, Zwischenzusammenfassungen und Sicherung von Teilergebnissen, Vorausblicke auf den Unterrichtsverlauf, Arbeitsrückschau, geordnete Darstellung der (Teil-)Inhalte in Form von Grafiken, Tafelbildern etc., explizite Verbindung von Teilergebnissen.



### Klarheit

Lernen ist ein Prozess der Informationsverarbeitung. Damit die notwendigen Informationen von SchülerInnen auch verstanden werden können, müssen sie mit hinreichender Klarheit formuliert sein. In Anlehnung an Helmke (2007, 2012) lassen sich vier Komponenten von Klarheit unterscheiden: *akustische* (Verstehbarkeit), *sprachliche* (Prägnanz), *inhaltliche* (Kohärenz) und *fachliche* (Korrektheit). Die akustische Klarheit stellt dabei die (physikalisch notwendige) Ausgangsbedingung für die drei weiteren Komponenten dar, welche sich auf die *inhaltliche Verständlichkeit* von Aussagen beziehen.

Um überprüfen zu können, ob die inhaltliche Verständlichkeit erreicht und somit die für sie charakteristischen drei Komponenten umgesetzt wurden, benennt Helmke (2012, S.191) folgende Indikatoren:

- 1) für *sprachliche Klarheit*: Vermeidung vager Ausdrücke; grammatische Korrektheit; Verwendung einfacher statt komplizierter Satzstrukturen; Vermeidung von Füll- und Verlegenheitswörtern
- 2) für *inhaltliche Klarheit*: eindeutige Formulierung von Aufgabenstellungen; Vermeidung redundanter Aussagen; Folgerichtigkeit der Aussagen; Ordnung und Gliederung der Aussagen; Unterscheidung von Wesentlichem und Unwesentlichem; hohes Maß an inhaltlicher Orientierung der SchülerInnen beim Angehen und Lösen von Aufgaben und beim Darstellen der Ergebnisse; inhaltliche Kohärenz vs. Sprunghaftigkeit
- 3) für *fachliche Korrektheit*: präziser Gebrauch von (Fach-)Begriffen.

### Exemplarizität (Einheit von Konkretem und Abstraktem)

In den 1960er Jahren wurde das Prinzip des Exemplarischen Lehrens und Lernens entwickelt. Den Grundgedanken des Exemplarischen Lehrens und Lernens umschrieb Klafki damals so: „Bildendes Lernen, das die Selbständigkeit des Lernenden fördert, also zu weiterwirkenden Erkenntnissen, Fähigkeiten, Einstellungen führt ... wird nicht durch reproduktive Übernahme möglichst vieler Einzelkenntnisse, -fähigkeiten und -fertigkeiten gewonnen, sondern dadurch, daß sich der Lernende an einer begrenzten Zahl von ausgewählten Beispielen [Exempeln] aktiv allgemeine, genauer: mehr oder minder weitreichend verallgemeinerbare Kenntnisse, Fähigkeiten, Einstellungen erarbeitet [...]“ (Klafki, 2007, S. 143f.). Im Sinne dieses Prinzips zielt lern-wirksamer Unterricht nicht in erster Linie auf den Erwerb positiven Wissens (im Sinne von Faktenwissen), sondern auf ein tiefes Verständnis der Sache (s. z.B. Klafki, 1961, 2007).

Die in den 1960er Jahren diagnostizierten Probleme (oberflächliches Wissen, kein wirkliches Verständnis der Sache usw.) scheinen jedoch von langfristiger Art zu sein, wie z.B. die Ergebnisse der *PISA-Studie 2000* (*Programme for International Student Assessment*, s. z.B. Baumert, Stanat, & Demmrich, 2001) vermuten lassen. Diese Studie bescheinigte deutschen SchülerInnen u.a. erhebliche Schwierigkeiten „im Bereich des naturwissenschaftlichen Verständnisses und bei der Anwendung ihres Wissens“ (Prenzel, Rost, Senkbeil, Häußler, & Klopp, 2001, S. 244). Deshalb gelte es, die „erkennbare Neigung zum fragend-entwickelnden und fachsystematisch orientierten Unterricht zu überwinden und durch Anwendungsbezug, Problemorientierung sowie Betonung mentaler Modelle das Interesse an den Naturwissenschaften und die Entwicklung eines tiefer gehenden Verständnisses und flexibel anwendbaren Wissens zu fördern“ (ebd., S. 245).

Die zu lernende „Sache“ wird aber nicht von den SchülerInnen abgebildet, sondern ist das Ergebnis eines Konstruktionsprozesses, durch den bestimmte Aspekte (Elemente) in spezieller Weise miteinander verknüpft werden. Daher müssen SchülerInnen die zu lernenden, von der Erwachsenenwelt (bzw. der Lehrperson) *vor-erfundene* Sache für sich neu *nach-konstruieren*.

Zur Unterstützung dieses Konstruktionsprozesses haben sich Beispiele bewährt (s. z.B. Durkin & Rittle-Johnson, 2012; Renkl et al., 2002; Wagenschein, 1968). Mit ihnen ist es möglich, eine Verbindung zwischen lebensweltlichen Phänomenen, bestehenden Vorwissensstrukturen und den neu zu erarbeitenden kognitiven Strukturen herzustellen. Um solche Verbindungen knüpfen zu können, ist jedoch wichtig, prägnante („passende“) Beispiele zu wählen. Ungeeignete Beispiele führen nämlich zu eher losen Assoziationen, Unverständnis, inhaltlichen Abschweifungen, mangelnder Transparenz usw. Deshalb muss die Lehrperson das zum Einsatz kommende Beispiel sorgfältig auswählen und überprüfen, ob in ihm die betreffenden Elemente der zu lernenden (kognitiven) Struktur erkennbar werden. Ist diese Forderung eingelöst, dann besteht Strukturisomorphie zwischen Beispiel und kognitiver Struktur bzw. im Sinne Klingbergs (1978) die „Einheit von Konkretem und Abstraktem“.

Exemplarisches Lehren und Lernen stellt sich somit als ein Dreischritt dar: Im ersten Schritt muss die Lehrperson ein konkretes Beispiel, einen prägnanten Fall, einen typischen Gegenstand usw. auswählen. Anhand dieser konkreten Gegebenheit sollen die SchülerInnen in einem zweiten Schritt wesentliche, d.h. allgemeine bzw. verallgemeinerbare Erkenntnisse gewinnen, die sie dann im dritten Schritt auf strukturgleiche Sachverhalte transferieren können.

Der wesentliche Indikator für Exemplarisches Lehren und Lernen stellt das Postulat der *Strukturisomorphie* dar, also die prägnante Passung von Beispiel und zu erarbeitender neuer kognitiver Struktur.

### **Lebensweltbezug des Themas**

SchülerInnen werden beim Lernen unterstützt, wenn sich die zu lernenden Inhalte auf deren Lebenswelt bzw. auf bedeutsame Kontexte ihrer Lebenswelt beziehen. Lehrpersonen haben demnach die Aufgabe, die zu lernenden fachlichen Inhalte in entsprechende Situationen einzubetten, damit einerseits das Interesse der SchülerInnen geweckt und andererseits dem Aufbau „trägen Wissens“ (s. z.B. Gruber & Renkl, 2000; Gruber, Mandl, & Renkl, 2000; Renkl, 1996, 1998; Renkl & Nückles, 2006) entgegengewirkt wird (s. z.B. Girwidz & Berger, 2008; Labudde, 1997; Mikelskis-Seifert & Duit, 2010; Muckenfuß, 1995; Parchmann et al., 2000).

Damit dieses „situiertere Lernen“ (Mandl & Kopp, 2005) nicht dazu führt, dass die jeweiligen Inhalte nur in den konkreten Situationen abgerufen werden können, sondern *flexibel* anwendbar sind, muss das sinnstiftende Lernen mit einer gewissen Variation der Kontexte einhergehen. Diese Variationen sollten eine Verbindung zur Phasierung des Unterrichts aufweisen, sodass sich bspw. die Einstiegs- und Erarbeitungsphase des neu zu erlernenden Wissens stimmig auf *ein und denselben* Kontext beziehen. Erst in einer anschließenden Vertiefungsphase sollten unterschiedliche Anwendungskontexte aufgezeigt werden, um das gelernte Wissen zu *dekontextualisieren* (s. Renkl, 2001).

In Bezug auf die Phasen des Unterrichts könnten mögliche Indikatoren für den Lebensweltbezug sein: geeignete *lebensweltliche Kontexte* für den *Einstieg*, *Erarbeitung* und *Anwendung* des Gelernten bzw. dessen *Transfer*.

### **Verknüpfung mit dem Vorwissen**

Das Vorwissen stellt einen, wenn nicht den, zentralen Prädiktor für den Lernerfolg der Lernenden dar (Ausubel et al., 1980, S. 65; Walberg & Paik, 2000, S. 13; Vosniadou, 2001, S. 12, 18; Wang et al., 1993), denn die zur Bildung einer neuen kognitiven Struktur notwendigen Elemente stammen zum großen Teil aus dem Vorwissen des Lernenden. Zusätzlich zum bisherigen Vorwissen müssen allerdings auch immer neue Informationen aufgenommen, verarbeitet und in geeigneter Weise in das bisherige Vorwissen integriert werden (Ausubel et al., 1980, S. 65; Reynolds, 1992). Lernaufgaben sollten daher so konzipiert sein, dass der Lernende die neue, zu erwerbende Struktur in Beziehung setzen kann zu jenen kognitiven Strukturen, über die er bereits verfügt. Ist das zum Erwerb neuer kognitiver Strukturen erforderliche individuelle Vorwissen defizitär oder nicht konsolidiert und deshalb nicht schnell und flexibel abrufbar, liegen ungünstige Bedingungen für das weitere Lernen vor (Bolhuis, 2003, S. 331; Hattie, 2009).

Mögliche Indikatoren für das Kriterium „Verknüpfung mit dem Vorwissen“ sind: Wiederholungsphase zu Beginn des Unterrichts, durch die das notwendige Vorwissen reaktiviert wird; Erzeugung von Widersprüchen zum bereits vorhandenen Wissen; direkte und indirekte Impulse im Unterrichtsverlauf, durch die an bereits Gelerntes erinnert wird; Vergleich der Relevanz der neuen Erkenntnisse mit den vorherigen.

### **Aktivierung/Motivierung zur geistigen Selbsttätigkeit (kognitive Aktivierung)**

In Abhängigkeit von den Zielsetzungen des Unterrichts muss die Lehrperson eine Lernumgebung schaffen, „in which students maximize the possibility of executing the cognitive activities necessary for building knowledge“ (Seidel & Shavelson, 2007, S. 458). Hierbei wird Lernen als ein *aktiver, mentaler Konstruktionsprozess* verstanden (s. z.B. Baumert & Köller, 2000, S. 273f.), durch den kognitive Strukturen auf konkrete Realität projiziert werden. Zum Beispiel sind Naturphänomene zwar sinnlich erfahrbare, aber verstanden, d.h. physikalisch auf den Begriff gebracht, werden sie erst durch spezifische Denkleistungen, durch die das Phänomen physikalisch konstruiert wird.

Solche Denkleistungen können aus *kognitionspsychologischer* Sicht mit der Anwendung bestehender und der Bildung neuer kognitiver Strukturen gleichgesetzt werden. Solche kognitiven Strukturen können in einem spezifischen Sinne auch als Handlungen, Operationen oder Begriffe verstanden werden (s. Aebli, 1993a), denen unterschiedliche Funktionen für die Erkenntnis- und Denkprozesse zukommen (s. dazu z.B. Gagel, 2000, S. 224ff.; Seiler, 1973, S. 12f.; Kleickmann, 2012): Kognitive Strukturen sind insofern *selektiv*, da nur spezifische Relationen zwischen Elementen (z.B. Begriffselementen) hergestellt werden (können). Durch diese Selektivität bestimmen die kognitiven Strukturen, wie und was wahrgenommen, welches Problem gesehen, wie es gelöst wird usw. Gleichzeitig besitzt die Selektivität aber einen *ordnungs-* und *bedeutungsstiftenden* Sinn, denn erst die zur Verfügung stehenden Kategorien ermöglichen eine strukturierte Informationsaufnahme und -verarbeitung.

Die im Unterricht zu erwerbenden kognitiven Strukturen stehen für kulturelle Objektivationen, die oft erst durch einen langen Entwicklungsprozess, der ggf. durch Irrtümer und

Kontingenzen gekennzeichnet war, zustande gekommen sind. Daher können im Unterricht, insbesondere mit Blick auf die zeitlichen Ressourcen sowie auf das kognitive Potential der SchülerInnen, nicht sämtliche kulturellen Errungenschaften selbstständig (ohne fremde Hilfe) erfunden werden. Aus diesem Grund müssen im unterrichtlichen Kontext das Lernen und Lehren als eine Einheit angesehen werden, bei der das *Lernen* der SchülerInnen dadurch bestimmt ist, dass sie neue kognitive Strukturen nicht (völlig) selbstständig bilden *können*, sie aber selbst bilden *sollen*. Das *Lehren* bestünde dann demnach in der *Bereitstellung professioneller Unterstützungsmaßnahmen*. Diese Hilfs- bzw. Unterstützungsmaßnahmen können im konstruktivistischen Verständnis unter dem Begriff der *kognitiven Aktivierung*, in didaktisch-methodischer Hinsicht eher unter dem der *Aktivierung/Motivierung zur geistigen Selbsttätigkeit* subsumiert werden.

Mittels der Aktivierung zur geistigen Selbsttätigkeit sollen in einem allgemeinen Sinne kognitive *Strukturdefizite* beseitigt werden. Kognitive Strukturen gelten dann als „defizitär“, (a) wenn zu ihrer Bildung notwendige Elemente fehlen, (b) wenn die Beziehungen zwischen den Elementen inhaltlich unzutreffend sind oder (c) wenn die in Frage kommenden Elemente nicht als „Ganzes“, als in sich stimmige, sinnbildende Struktur gesehen werden. Im Unterricht sollten diese Punkte berücksichtigt und produktiv genutzt werden. Es käme also darauf an, (a) die notwendigen Elemente vollständig zu präsentieren, (b) korrekte inhaltlich spezifische Beziehungen zwischen ihnen herzustellen, sowie (c) auf die angemessene Stiftung räumlicher und zeitlicher Nähe zu achten, damit die Bildung der betreffenden Struktur als Ganzes einsichtig wird. Die folgenden Indikatoren benennen eine Auswahl verschiedener Möglichkeiten zur Aktivierung/Motivierung zur geistigen Selbsttätigkeit:

- Zu (a): Erinnerung an bereits erworbene Wissens Elemente; explizite und implizite Einführung von weiteren Wissens Elementen in den Unterrichtsdiskurs; Unvollständiges präsentieren; Wissenslücken benennen lassen; Problemstellungen formulieren lassen.
- Zu (b): Widersprüche (kognitive Dissonanzen) erzeugen; Situationen schaffen, in denen die SchülerInnen die Unzulänglichkeiten ihrer kognitiven Strukturen registrieren; Verbesserung mangelhafter Lösungen; Vergleiche auf Gemeinsamkeiten und Unterschiede von Aspekten; Querverbindungen herstellen lassen; „Feinheiten“ entdecken lassen.
- Zu (c): Ordnung in „ungeordnete“ Sachverhalte bringen; Prüfung neuer Ideen; Suche nach Regelhaftem, Typischem, Strukturellem, Gesetzmäßigem oder nach Form-Inhalt-Relationen; grafische Darstellung von Zusammenhängen; Lernergebnisse in eigenen Worten zusammenfassen lassen; Begriffe ausschärfen lassen.

### **Diskursivität des Unterrichtsprozesses (Reflexion von Inhalt und Methodik)**

Die bisher genannten Prozessmerkmale stellen Handlungsoptionen für Lehrpersonen dar, durch die sie den Lernprozess der SchülerInnen unterstützen können. Sie sind keine technischen „Instrumente“, mit deren Hilfe die Lehrperson etwas vermitteln kann, sondern handlungsleitende Prinzipien, die einen Kommunikationsprozess zwischen Lehrpersonen und SchülerInnen in Gang bringen und halten sollen, um *gemeinsame* Sinnstrukturen aufzubauen. „Gemeinsam“ ist der konstituierte Sinn, weil er versprachlicht vorliegt und für alle Beteiligten fassbar ist. Daraus folgt aber nicht, dass der konstruierte Sinn allseitige

Zustimmung findet (etwa bei normativ aufgeladenen Themen, Diskussionen, der Interpretation fiktionaler Literatur, usw.), gleichwohl dies durchaus der Fall sein kann (z.B. bei der Konstruktion physikalischer Begriffe, die experimentellen Daten standhalten müssen). Ein Unterricht, der dieser Prämisse folgt (Kommunikation zum Aufbau gemeinsamer Sinnkonstruktionen), setzt nicht „auf passives, nachvollziehendes Lernen, auf die Rezeption von Vorgeordnetem“ (Schmidt-Wulffen, 1989, S. 22) und die unkritische Wissensübertragung. Diskursiver Unterricht meint vielmehr: „Von einer Vorstellung zur anderen voranzuschreiten, durch Erörtern und Argumentieren. Dabei sollte von den geläufigen Auffassungen ausgegangen werden. Deren Begrenztheit gilt es zu verdeutlichen, deren Vorläufigkeit und Unvollkommenheit bewußtzumachen. Neue Informationen, ein neuer Zugang stellt das bisher Angenommene in Frage, treibt den Lernprozeß voran.“ (ebd., S. 25) Ein solcher Unterricht setzt auf *Einsicht* seitens der SchülerInnen, sodass im unterrichtlichen Diskurs Argumente verhandelt werden und auch weiterhin verhandelbar bleiben (s. Cohors-Fresenborg & Kaune, 2003, 2007; Reiss, 2002; Schmidt-Wulffen, 1989; Vielhaber, 2001). Vielhaber (2001) meint dazu: „Ein Unterricht, der diskursförmigen Verständigungen eine Plattform gibt, um genau das, nämlich die *Einsichtsfähigkeit*, zu ermöglichen, eröffnet allen Teilnehmern die Möglichkeit, in Rede und Gegenrede, in Frage und Antwort an der Kommunikation teilzunehmen, immer unter der Voraussetzung, daß die diskursiven Verfahrensregeln akzeptiert wurden.“ (Vielhaber, 2001, S. 92) Der Lehrperson kommt dabei die Aufgabe zu, diesen Diskurs zu initiieren und aufrecht zu erhalten und dabei eine Balance zwischen hinreichender Offenheit und notwendiger Geschlossenheit der Kommunikation zu gewährleisten.

Derartige Argumente beziehen sich aber nicht nur auf die inhaltliche Progression des Unterrichts, sondern auch auf das methodische Vorgehen im Unterricht. So geht es beispielsweise im Physikunterricht nicht nur darum, einen Begriff, wie etwa den des „Kräfteparallelogramm“ kennenzulernen. Auch die genetische Entwicklung dieses Begriffes ist diskursiv zu verhandeln, sodass SchülerInnen die Möglichkeit bekommen, zunächst Hypothesen zu bilden und Überlegungen anzustellen, ob sich eine resultierende Kraft in Teilkräfte zerlegen lässt. Darüber hinaus müssten die SchülerInnen überlegen, wie diese Teilkräfte im Verhältnis zur resultierenden Kraft stehen und wie diese Größen gemessen werden können.

Als mögliche Indikatoren für die Diskursivität des Unterrichtsprozesses könnten herangezogen werden: Rede und Gegenrede zulassen; das Infragestellen von Argumenten provozieren bzw. unterstützen; Begründungen und Belege für Argumente einfordern; Zulassen alternativer Denk- und Argumentationsmuster; Bemühen um Konsens in der Auswahl der besten Argumente; Redebeiträge von SchülerInnen in weiterführende Impulse umformulieren; Einbringen und Bewerten weiterer Arbeitsschritte; Diskussion von Ideen und Hypothesen; Überlegungen, wie man das Thema angehen könnte; explizite Stellungnahme von SchülerInnen zur subjektiven Bedeutung des Themas; Begründungen für die Angemessenheit des methodischen Vorgehens einfordern; SchülerInnen auffordern, ihre Lernschwierigkeiten zu explizieren; gemeinsam nach Ursachen für Verständnisschwierigkeiten suchen.

### **Individualisierung/Differenzierung**

Die Lernvoraussetzungen und Leistungsunterschiede zwischen SchülerInnen können innerhalb einer Lerngruppe bzw. eines Klassenverbandes deutlich variieren. Daher muss die

Lehrperson geeignete Maßnahmen ergreifen, um dieser Heterogenität gerecht zu werden (s. z.B. Bönsch, 2011; Good & Brophy, 2000; Paradies & Linser, 2010; Rossbach & Wellenreuther, 2002; Schittko, 1984; Vollstädt, 1997; Walberg & Paik, 2000; Waxman et al., 1985; Wiater, 2008).

Zur Auswahl solcher Maßnahmen haben Klafki und Stöcker (in Klafki, 2007, S. 188) eine Matrix entwickelt, die ein weites Spektrum innerer Differenzierungsmöglichkeiten<sup>69</sup> beinhaltet. Diese Matrix ist in drei Dimensionen gegliedert, die jeweils verschiedene Differenzierungsoptionen umfassen:

- Dimension A: Unterrichtsphasen (Differenzierung in Phasen Aufgabenstellung bzw. -entwicklung; Erarbeitung; Festigung; Anwendung/Transfer),
- Dimension B: Differenzierungsaspekte im Hinblick auf den Schüler (Stoffumfang/ Zeitaufwand; Komplexitätsgrad; Anzahl der notwendigen Durchgänge; Notwendigkeit direkter Hilfe/Grad der Selbstständigkeit; Art der inhaltlichen oder methodischen Zugänge/der Vorerfahrungen; Kooperationsfähigkeit),
- Dimension C: Aneignungs- und Handlungsebenen (konkrete Aneignungs- bzw. Handlungsebene; explizit-sprachliche Aneignungs- bzw. Handlungsebene; rein gedankliche Aneignungs- bzw. Handlungsebene).

Die von Klafki und Stöcker genannten und in Klammern angeführten Aspekte stellen zugleich mögliche Indikatoren für die Analyse von Unterricht dar, die wie folgt ausformuliert werden können: unterschiedliche Zugänge zum Thema wählen; verschiedene Grade der Komplexität zulassen; zusätzliche Aktivitäten anbieten; bestimmten SchülerInnen mehr Zeit zur Bearbeitung von Aufgaben zugestehen; Lerngruppen flexibel einrichten; individuelle Hilfen bei der Aufgabenbearbeitung geben.

### Ergebnissicherung/Erfolgskontrolle

Ergebnissicherung und Erfolgskontrolle sollen Aufschluss darüber geben, ob die Lernziele einer Unterrichtsstunde erreicht wurden. Das bedeutet: Wenn der Unterricht als intendierte Lehre an Lernziele gebunden ist, dann muss (spätestens) zum Ende einer Unterrichtsstunde überprüft werden, ob die gesetzten Lernziele erreicht wurden.

Dabei zielt die *Ergebnissicherung* einerseits auf den Abgleich zwischen den Lehr- und Lernzielen und andererseits auf die Sicherung der erreichten Lernergebnisse (s. z.B. Helmke & Schrader, 2009, S. 710; Klingberg, 1978, S. 272; Meyer, 1987, S. 165; Voss & Kunter, 2011). Für die SchülerInnen bildet die Ergebnissicherung zugleich die *Erfolgskontrolle*, deren Funktion im Feedback auf bereits Erreichtes und möglicherweise in Hinweisen auf noch fehlendes Wissen liegt (Kyriakides et al., 2013, S. 145; Walberg & Paik, 2000, S. 17; Reynolds, 1992).<sup>70</sup> Darüber hinaus gibt die Erfolgskontrolle der Lehrperson Aufschluss über die *Effektivität* ihrer Unterrichtsplanung.

<sup>69</sup> Der Begriff „innere Differenzierung“ wird auch synonym mit Binnendifferenzierung verwendet und umfasst „alle Maßnahmen, die innerhalb einer Lerngruppe (Klasse) dazu beitragen, den unterschiedlichen Fähigkeiten, Fertigkeiten, Interessen und Lernbedürfnissen der Schüler zu entsprechen und deren Lernprozesse in optimaler Weise zu unterstützen“ (Vollstädt, 1997, S. 37).

<sup>70</sup> Damit der Lernprozess der SchülerInnen unterstützt wird, sollte die Lehrperson ihnen Feedback in Form lernrelevanter Informationen geben. Diese Informationen können zum einen am Ende einer Lerneinheit, in Form einer *summativen*, abschließenden Bewertung der Leistungsergebnisse verkündet werden; wichtiger erscheint aber eine *formative Evaluation*, d.h. den gesamten Unterrichtsprozess durchziehende Rückmeldung über den Ist-Soll-Stand (Kyriakides et al., 2013, S. 146). Solche formativen Rückmeldungen sind besonders dann leistungs- und motivationsfördernd, wenn sie sich auf die Bearbeitung einzelner *Aufgaben* (z.B. richtige



Mögliche Indikatoren für die Ergebnissicherung bzw. Erfolgskontrolle sind: explizite Besprechung des Ergebnisses; Verbalisieren des Lernzuwachses; Zusammenfassung des Wesentlichen; (abschließende) Gelegenheit, aus Fehlern zu lernen; hinreichende Schülerzentrierung; Blick auf das im Verlauf des Unterrichts genetisch erstellte, in der Schlussphase als Gesamtbild vorliegende Tafelbild; Gelegenheit, Fragen zum Unterrichtsverlauf und -ergebnis zu stellen.

### Anwendung/Transfer

Die erworbenen neuen bzw. modifizierten kognitiven Strukturen stellen (geistige) Werkzeuge zur Erschließung und zum Verständnis weiterer Situationen dar. Für SchülerInnen stellt es jedoch eine entsprechende Herausforderung dar, das Gelernte anzuwenden bzw. auf neue Situationen zu transferieren (Kyriakides et al., 2013, S. 145). In Anlehnung an die bereits erwähnte Strukturisomorphie, verstanden als Passung von Beispiel und kognitiver Struktur (s. das Merkmal *Exemplarizität*, S. 78f.), stellt das konkrete Beispiel nach Renkl (2001) jeweils die „Oberfläche“ dar. Aus diesen Oberflächenmerkmalen müssen sowohl wesentliche Elemente (die kognitive Struktur) herausgezogen als auch eine ganze Reihe lösungsirrelevanter Merkmale eliminiert werden, damit die Tiefenstrukturen der jeweiligen Sachverhalte erkennbar (erfassbar) werden (Chi & VanLehn, 2012; Renkl et al., 2009; Schwartz et al., 2011).

Damit die Anwendungs- und Transfermöglichkeiten dem Leistungsstand der SchülerInnen gerecht werden, sollten nach Renkl die zu lösenden Aufgaben zunächst nur geringfügig und dann stärker in ihren Oberflächenmerkmalen variiert werden. Erst wenn solche Aufgaben mit hinreichender Sicherheit gelöst werden können, sollten Aufgaben gewählt werden, für deren Lösung eine Änderung bestehender kognitiver Strukturen erforderlich sind. Dabei geht der Unterrichtsprozess in die Erarbeitung neuer Strukturen, im Sinne eines problemlösenden Aufbaus (Aebli, 1993a), über.

Als mögliche Indikatoren für die Analyse können genutzt werden: (steigende) Variation der Oberflächenmerkmale bei Aufgaben mit identischer Struktur; Passung der erlernten Struktur auf die Anwendungsaufgaben; Entwicklung neuer Problemstellungen aus entsprechenden Anwendungskontexten.

### Lernatmosphäre

Das Modell der Struktur- und Prozessqualität wird gerahmt von der *Lernatmosphäre* (s. Abb. 3, S. 24). In der Forschungsliteratur wird die Lernatmosphäre meistens unter dem Begriff „lernförderliches Klima“ subsumiert, mit dem eine Lernumgebung gemeint ist, „in der das Lernen der Schülerinnen und Schüler erleichtert, begünstigt oder auf andere Weise positiv beeinflusst wird.“ (Helmke, 2012, S. 226) Brophy (2000) führt dieses „supportive classroom climate“ als erstes seiner zusammengetragenen Prinzipien effektiven Lehrens an (s. Brophy 2000, S. 8f.):

To create a climate for moulding their students into a cohesive and supportive learning community, teachers need to display personal attributes that will make them effective as models and socializers: a cheerful disposition, friendliness, emotional maturity, sincerity, and caring about

---

und falsche Lösungen kenntlich machen) beziehen, Auskunft über den Fortschritt des bisherigen *Lernprozesses* geben und dem Schüler Feedback über *selbstregulatorische* (self-regulated) Prozesse liefert, aus denen er Rückschlüsse auf die Selbstwirksamkeit seines Handelns und Denkens ziehen kann (Brophy, 2000; Fraser et al., 1987; Hattie, 2009; Hattie & Wollenschläger, 2014).



students as individuals as well as learners. The teacher displays concern and affection for students, is attentive to their needs and emotions, and socializes them to display these same characteristics in their interactions with one another.  
(Brophy, 2000, S. 8)

Mit dieser Aussage Brophys werden vor allem die Unterstützung und die Herstellung eines lernförderlichen Klimas seitens der Lehrperson angesprochen (s. Helmke, 2012). Lipowsky gibt aber zu bedenken, dass sich aus empirischer Sicht nur wenige „direkte Effekte der affektiven Qualität der Lehrer-Schüler-Beziehung auf den Lernerfolg der Lernenden“ (Lipowsky, 2015, S. 92) nachweisen lassen. Es gäbe aber *indirekte* Effekte des Unterrichtsklimas auf den Lernerfolg:

In Klassen mit einem positiv ausgeprägten Klima, das von gegenseitiger Wertschätzung und Respekt geprägt ist, fühlen sich die Lernenden wohler. Sie gehen gerne in die Schule, haben positivere Beziehungen zu Mitlernenden und zu ihrer Lehrperson, erleben sich stärker sozial eingebunden und dazugehörig, strengen sich mehr an und entwickeln höhere Motivation für das Lernen. Das aktive Engagement und die höhere Motivation wirken sich dann wiederum positiv auf den Lernerfolg aus.

(Lipowsky, 2015, S. 92)

Eine wichtige Voraussetzung für ein lernförderliches Klima stellt die klare *Trennung* von Lern- und Leistungssituationen durch die Lehrperson dar. Die Lehrperson sollte die in vorangehenden Lernsituationen getätigten Schüleräußerungen ernstnehmen und mit ihnen produktiv umzugehen wissen. Das Aufgreifen der Beiträge dient dazu, dass die SchülerInnen eine gewisse Selbstwirksamkeit erfahren und dadurch sehen, dass die Beteiligung am Unterrichtsgeschehen und der Austausch mit den Mitschülern förderlich für das eigene Lernen sind.

Weitere Indikatoren für die Realisierung eines lernförderlichen Klimas sind (s. z.B. Brophy, 2000; Gruehn, 2000; Helmke, 2006, 2012; Klieme et al., 2006; Jahn et al., 2011; Meyer, 2003, 2013; Seidel, Prenzel, Rimmele, Schwindt et al., 2006): das Auftreten der Lehrperson ist durch Freundlichkeit, Wertschätzung, Unterstützung und Humor geprägt; gegenseitiger Respekt aller im Klassenzimmer anwesenden Personen; die Lehrperson vermeidet und unterbindet sozial vergleichende Beurteilungen und Beschämung; Abbau hemmender Leistungsangst; konstruktiver Umgang mit Fehlern.

### 3.1.3 Kategorien für didaktisch-methodisches Wissen

Diese, im vorangegangenen Teil, theoretisch begründeten 12 Prozessmerkmale, nutze ich als Kategorien im Rahmen der Qualitativen Inhaltsanalyse meiner Untersuchung. In diesem Teil werden diese Kategorien nun in tabellarischer Form vorgestellt und, in Anlehnung an die Forderungen aus der einschlägigen Methodenliteratur (s. z.B. Kuckartz, 2016; Lissmann, 2008; Mayring, 2015), mittels präziser *Definition*, *Beschreibung* und je einem aussagekräftigen *Ankerbeispiel* operationalisiert.

Zielorientierung

Tab. 11: Kategorie Zielorientierung

Kategorie	Definition/Beschreibung/Ankerbeispiel
Zielorientierung (Kürzel: ZO)	<p><i>Definition:</i> Zielorientierung bedeutet, dass der Verlauf von Unterricht an einem übergreifenden Ziel (und ggf. an weiteren Teilzielen) ausgerichtet ist.</p> <p><i>Beschreibung:</i> Unterricht als intendierte Lehre ist an Lernziele gebunden. Diese sollten möglichst klar definiert und in Teilziele untergliedert sein. Im Verlauf des Unterrichtsprozesses kommt es darauf an, dass die SchülerInnen die Lehrziele als eigene Lernziele übernehmen und ihnen dadurch der Unterrichtsprozess transparent wird.</p> <p><i>Ankerbeispiel:</i> „Er stellt keinen weiteren Zusammenhang her. Also wenn ich in die Mienen der SchülerInnen bei seinen langen Vorträgen sehe, die wissen gar nicht, worum es geht. Wo will er denn eigentlich hin? Ich sage immer, das ist wie Stochern im Nebel. Der Lehrer gibt einen Impuls, die SchülerInnen wissen, sie sollen jetzt etwas sagen, aber sie haben eigentlich kein Ziel vor Augen. Sie wissen gar nicht, wo es hingehen soll.“</p>

Strukturierung

Tab. 12: Kategorie Strukturierung

Kategorie	Definition/Beschreibung/Ankerbeispiel
Strukturierung (Kürzel: S)	<p><i>Definition:</i> Strukturierter Unterricht ist gekennzeichnet durch aufeinander folgende Phasen, in denen schrittweise inhaltliche Komplexität erzeugt wird.</p> <p><i>Beschreibung:</i> Die im Unterricht anzueignenden Sachverhalte sind relativ komplex, sodass es notwendig ist, sie schrittweise zu erarbeiten. Dies geschieht dadurch, dass der gesamte Lernprozess in mehrere, inhaltlich aufeinander bezogene Teilprozesse gegliedert wird. Auf diese Weise wird eine stimmige Verknüpfung und Aufeinanderfolge von Unterrichtsphasen realisiert.</p> <p><i>Ankerbeispiel:</i> „Er hat sicher eine vernünftige Planung gemacht, aber bei der Strukturierung der Stunde hat er die Abhängigkeit zwischen den einzelnen Phasen nicht berücksichtigt, wodurch immer wieder Brüche auftreten und die vorherigen Phasen nicht berücksichtigt werden. Diese Planungsproblematik hat sich eigentlich durch die ganze Stunde gezogen.“</p>

## Klarheit

**Tab. 13:** Kategorie Klarheit

Kategorie	Definition/Beschreibung/Ankerbeispiel
Klarheit (Kürzel: K)	<p><i>Definition:</i> Klarheit ist im Unterricht gegeben, wenn Informationen so formuliert sind, dass SchülerInnen diese verstehen können.</p> <p><i>Beschreibung:</i> Lernen ist ein Prozess der Informationsverarbeitung. Damit die notwendigen Informationen von SchülerInnen verstanden werden können, müssen sie mit hinreichender Klarheit formuliert sein. Unter Klarheit wird hier im engeren Sinne sprachliche Prägnanz verstanden. Sie ist gekennzeichnet durch einen präzisen Ausdruck, grammatische Korrektheit, einfache statt komplizierte Satzstrukturen und angemessene Informationsfülle.</p> <p><i>Ankerbeispiel:</i> „... das Erklären selbst ist, und das hat man in den Schülergesichtern gesehen, ist ihm nicht gelungen. Er hat sie zwischendurch abgehängt, hat sie erstens nicht da abgeholt, wo sie stehen, und dann durch ein Übermaß an Sprache auch abgehängt.“</p>

## Exemplarizität (Einheit von Konkretem und Abstraktem)

**Tab. 14:** Kategorie Exemplarizität (Einheit von Konkretem und Abstraktem)

Kategorie	Definition/Beschreibung/Ankerbeispiel
Exemplarizität (Einheit von Konkretem und Abstraktem) (Kürzel: Exe)	<p><i>Definition:</i> Das Prinzip der Exemplarizität wird im Unterricht durch die präzise Abstimmung von konkreten Beispielen auf die neu zu erarbeitenden kognitiven Strukturen eingelöst.</p> <p><i>Beschreibung:</i> Exemplarisches Lehren und Lernen lässt sich durch einen Drei-Schritt kennzeichnen: Anhand eines konkreten Beispiels wird eine allgemeine Erkenntnis (kognitive Struktur) erarbeitet. Diese allgemeine Erkenntnis ermöglicht den SchülerInnen den Transfer auf weitere (strukturelle) Fälle. Entscheidend bei diesem Vorgehen ist, dass die konkreten Beispiele genau diejenigen strukturellen Momente enthalten, die für die zu erarbeitenden kognitiven Strukturen wesentlich sind (Strukturisomorphie). Ist dies der Fall, dann ist die Einheit von Konkretem und Abstraktem gegeben.</p> <p><i>Ankerbeispiel:</i> „Was mir relativ gut gefallen hat, ist, dass er mehrere Phänomene zusammengestellt hat, bei denen die Lichtbrechung eine Rolle spielt. Ob die ausgewählten Beispiele jetzt alle so furchtbar clever ausgewählt waren und tatsächlich so geeignet sind, um zu dem Phänomen hinzuführen, zu welchem er hinführen möchte, muss bezweifelt werden. Warum macht er das Problem nicht anschaulicher? Warum stellt er nicht einfach eine Schüssel Wasser hin, beleuchtet diese und lässt beobachten?“</p>

Lebensweltbezug des Themas

Tab. 15: Kategorie Lebensweltbezug des Themas

Kategorie	Definition/Beschreibung/Ankerbeispiel
Lebensweltbezug des Themas (Kürzel: LWB)	<p><i>Definition:</i> Lebensweltbezug des Themas zielt auf die Integration von konkreten Alltagserfahrungen der SchülerInnen in den Unterricht, um sinnstiftendes Lernen zu ermöglichen.</p> <p><i>Beschreibung:</i> Sinnstiftendes Lernen wird für SchülerInnen erleichtert, wenn sich fachliche Inhalte auf bedeutsame Kontexte aus der Lebenswelt beziehen. Denn durch diese Integration von Alltagserfahrung in den Unterricht wird nicht nur das Interesse der SchülerInnen geweckt, sondern auch dem Aufbau „trägen Wissens“ entgegengewirkt.</p> <p><i>Ankerbeispiel:</i> „Die Frage ist im Grunde elementar. Ich habe jetzt eine Gesetzmäßigkeit, was kann ich damit im Alltag anfangen, wo kann ich das anwenden. Im Fachbereich? In der Fachwissenschaft oder auch wirklich im Alltag? Da hätte man die SchülerInnen auch mal zu Wort kommen lassen müssen, welche Ideen haben die. Denn klar ist, was ich mir so aneigne, muss natürlich irgendwo auch für mich eine Bedeutung haben im Unterricht. Die werden aus der Stunde rausgehen und sich darüber keine Gedanken mehr machen.“</p>

Verknüpfung mit dem Vorwissen

Tab. 16: Kategorie Verknüpfung mit dem Vorwissen

Kategorie	Definition/Beschreibung/Ankerbeispiel
Verknüpfung mit dem Vorwissen (Kürzel: VW)	<p><i>Definition:</i> Die Verknüpfung mit dem Vorwissen meint die spezifische Beziehung, in die bestehende kognitive Strukturen der SchülerInnen mit neu zu erwerbenden Strukturen gesetzt werden.</p> <p><i>Beschreibung:</i> Um sinnvolles Lernen zu ermöglichen, muss die Lehrperson Maßnahmen ergreifen, durch die neue Wissenselemente in bestehende Wissensstrukturen seitens der Lernenden integriert werden können. Lernaufgaben müssen daher so konzipiert sein, dass das Vorwissen der Lernenden und die kognitiven Anforderungen, die bspw. durch Aufgaben repräsentiert werden, stringent aufeinander bezogen werden (König, Buchholtz, &amp; Dohmen, 2015).</p> <p><i>Ankerbeispiel:</i> „Also er bemüht sich hier, das Vorwissen der SchülerInnen abzugreifen, aber er lässt den SchülerInnen wenig Spielraum, um eigene Ideen zu entwickeln bzw. das aufzugreifen, was sie im Grunde schon an Vorwissen haben. Es wäre jetzt auch für mich interessant gewesen, ganz präzise zu wissen, welche Lernvoraussetzungen sind gegeben. In welcher Präzision sind die Begrifflichkeiten eingeführt? In der Ausbildung empfehlen wir, diese Phasen stärker zu trennen und den SchülerInnen an dieser Stelle unser Wissen, unser Vorwissen über Brechung, über optische Phänomene nochmal auffrischen lassen. Dazu schaffen wir eine Plattform.“</p>

**Aktivierung/Motivierung zur geistigen Selbsttätigkeit (kognitive Aktivierung)****Tab. 17:** Kategorie Aktivierung/Motivierung zur geistigen Selbsttätigkeit  
(kognitive Aktivierung)

Kategorie	Definition/Beschreibung/Ankerbeispiel
Aktivierung/Motivierung zur geistigen Selbsttätigkeit (kognitive Aktivierung) (Kürzel: AMgS)	<p><i>Definition:</i> Die Kategorie „Aktivierung/Motivierung zur geistigen Selbsttätigkeit“ bezieht sich auf die Aufgabe der Lehrperson, die Denkprozesse der SchülerInnen durch geeignete Maßnahmen zu initiieren und in ihrem Ablauf zu unterstützen.</p> <p><i>Beschreibung:</i> Lernen ist ein aktiver mentaler Konstruktionsprozess, bei dem es um den Aufbau neuer kognitiver Strukturen geht. Diese Strukturen können die SchülerInnen letztlich aber nur selbsttätig bilden. Die Aufgabe der Lehrperson besteht demnach darin, durch spezifische Lehr-Lernarrangements diese Selbsttätigkeit (kognitive Aktivierung) anzuregen. Dazu sollten SchülerInnen die Möglichkeit bekommen, den Prozess des Lernens selbstständig und eigenverantwortlich zu gestalten.</p> <p><i>Ankerbeispiel:</i> „Da zeigt sich wieder: Die dürfen nichts selbst entwickeln. Die SchülerInnen haben überhaupt keine Chance eigene Ideen einzubringen, sie müssen dem Gedanken des Lehrers folgen. Eine Offenheit fehlt vollständig. Das liegt daran, dass er die Lernerseite nicht genügend im Blick hat; er hat die Lehrseite im Blick, aber er muss in die Köpfe der SchülerInnen hineinschauen. Welche Vorstellungen existieren da? Das ist bei ihm nicht ausgeprägt genug.“</p>

Diskursivität des Unterrichtsprozesses (Reflexion von Inhalt und Methode)

Tab. 18: Kategorie Diskursivität des Unterrichtsprozesses (Reflexion von Inhalt und Methode)

Kategorie	Definition/Beschreibung/Ankerbeispiel
Diskursivität des Unterrichtsprozesses/ Reflexion von Inhalt und Methode (Kürzel: Disk)	<p><i>Definition:</i> Diskursivität bedeutet, dass die inhaltlichen und methodischen Aspekte des Unterrichtsprozesses nicht allein von der Lehrperson vorgegeben, sondern unter Beteiligung der SchülerInnen verhandelt werden bzw. verhandelbar bleiben.</p> <p><i>Beschreibung:</i> Unterricht ist kein reiner Vermittlungsprozess im Sinne unkritischer Übernahme bestehenden Wissens, sondern Unterricht wird als kommunikatives Geschehen aufgefasst, indem es um den Aufbau von gemeinsamen Sinnstrukturen geht. Diskursiv ist dieser Prozess dann angelegt, wenn sowohl die inhaltlichen als auch die methodischen Aspekte in der unterrichtlichen Kommunikation (prinzipiell) verhandelbar sind. Die Lehrperson hat dabei die Funktion, den Diskurs in Gang zu setzen und zu halten. Sie muss dabei eine Balance zwischen hinreichender Offenheit und notwendiger Geschlossenheit des Unterrichtsprozesses herstellen.</p> <p><i>Ankerbeispiel:</i> „Die Auswertung der Messergebnisse finde ich nicht gelungen. Die SchülerInnen haben Messungen vorgenommen und dann auf ihrem Blatt eingetragen. Normal wäre jetzt, dass die Lehrperson fragt, welche Ergebnisse die SchülerInnen herausbekommen haben, dass er vielleicht sogar die Chance nutzt, die SchülerInnen miteinander ins Gespräch kommen zu lassen und sie ihre Werte austauschen lässt, denn es kann ja auch Differenzen geben. Und dann müssten die SchülerInnen gemeinsam überlegen, ob sich aus den Daten ein Gesetz ableiten lässt. Aber er wertet die Messergebnisse für die SchülerInnen aus und so ist keine Aktivität mehr bei den SchülerInnen zu erkennen. Sie bekommen nicht die Chance, eigene Ideen oder Fragen zu produzieren, denn er gibt sich selbst die Antworten.“</p>

Individualisierung/Differenzierung

Tab. 19: Kategorie Individualisierung/Differenzierung

Kategorie	Definition/Beschreibung/Ankerbeispiel
Individualisierung/ Differenzierung (Kürzel: InDi)	<p><i>Definition:</i> Differenzierung und Individualisierung beschreiben alle Maßnahmen der Lehrperson, die auf die individuellen Lernvoraussetzungen der SchülerInnen zugeschnitten sind.</p> <p><i>Beschreibung:</i> Das Leistungspotential von SchülerInnen kann in einer Lerngruppe beträchtlich differieren. Insofern sind unterrichtliche Anforderungen und unterstützende Maßnahmen auf die individuelle Leistungsfähigkeit der SchülerInnen abzustimmen.</p> <p><i>Ankerbeispiel:</i> „Der Lehrer nimmt keine Rücksicht auf die SchülerInnen und ihre individuellen Voraussetzungen, die sie mitbringen. Aber er müsste sich zu Beginn der Stunde einen Überblick darüber verschaffen, was die Schüler können und dann, je nach Grad der Notwendigkeit, individuelle Hilfestellungen anbieten. Aber das macht er nicht.“</p>

## Ergebnissicherung/Erfolgskontrolle

**Tab. 20:** Kategorie Ergebnissicherung/Erfolgskontrolle

Kategorie	Definition/Beschreibung/Ankerbeispiel
Ergebnissicherung/ Erfolgskontrolle (Kürzel: EsEk)	<p><i>Definition:</i> Ergebnissicherung und Erfolgskontrolle sollen Aufschluss darüber geben, ob die Lernziele einer Unterrichtsstunde erreicht wurden.</p> <p><i>Beschreibung:</i> Wenn der Unterricht als intendierte Lehre an Lernziele gebunden ist, dann muss (spätestens) zum Ende einer Unterrichtsstunde überprüft werden, ob die gesetzten Lernziele erreicht wurden.</p> <p><i>Ergebnissicherung</i> zielt auf den Abgleich zwischen den Lehrzielen einerseits und erreichten Lernergebnissen andererseits. Für die SchülerInnen bildet sie zugleich die <i>Erfolgskontrolle</i>, deren Funktion im Feedback auf bereits erreichtem und möglicherweise noch fehlendem Wissen liegt.</p> <p><i>Ankerbeispiel:</i> „Die SchülerInnen haben offensichtlich gearbeitet, aber es findet keine Präsentationsphase statt. Keiner der SchülerInnen gibt auch nur einen Wert an, den er gemessen hat. Das ist ja eine interessante Phase, in der man als Lehrer sehen kann: Konnten die das Experiment durchführen? Haben die sauber gearbeitet? Haben die annähernd die richtigen Werte? Liegen die Werte auseinander? Da kann man ganz schnell auf einen Blick mitbekommen, wo die SchülerInnen stehen. Und die SchülerInnen könnten übrigens sehen, wenn sie jetzt wirklich gravierende Fehler gemacht hätten, woran das lag, und kämen noch entsprechend zur Einsicht. Diese Chancen werden hier überhaupt nicht genutzt.“</p>

## Anwendung/Transfer

**Tab. 21:** Kategorie Anwendung/Transfer

Kategorie	Definition/Beschreibung/Ankerbeispiel
Anwendung/Transfer (Kürzel: AT)	<p><i>Definition:</i> Anwendung und Transfer beschreiben die Übertragung erworbener kognitiver Strukturen auf (weitere) konkrete Sachverhalte.</p> <p><i>Beschreibung:</i> Neu erworbene kognitive Strukturen implizieren ein entsprechendes Transferpotential, das SchülerInnen eine vielfältige Anwendung dieser Erkenntnisse ermöglicht. Auf diese Weise werden die in einem spezifischen Kontext (z.B. an konkreten Beispielen) erworbenen Strukturen wieder dekontextualisiert.</p> <p><i>Ankerbeispiel:</i> „Die Fresnel-Linse, die hier als Anwendungsbeispiel zum Schluss der Stunde eingesetzt wird, stellt eine Überforderung der SchülerInnen dar, denn die können gar nicht nachvollziehen, wozu diese Linse gebraucht wird, da er nicht die entsprechenden Grundlagen legt. Für die SchülerInnen bleibt so vielleicht eine interessante Erfahrung, diese Linse mal gesehen zu haben, aber wieso und weshalb diese Linse so besonders ist, wird nicht deutlich.“</p>



Lernatmosphäre

Tab. 22: Kategorie Lernatmosphäre

Kategorie	Definition/Beschreibung/Ankerbeispiel
Lernatmosphäre (Kürzel: LA)	<p><i>Definition:</i></p> <p>Unter Lernatmosphäre wird die Qualität des Umgangs zwischen Lehrperson und SchülerInnen sowie den SchülerInnen untereinander verstanden.</p> <p><i>Beschreibung:</i></p> <p>Als Voraussetzung für ein lernförderliches Klima und für angstfreies Lernen muss Unterricht durch einen respektvollen und höflichen Umgang von Lehrperson und SchülerInnen, aber auch der SchülerInnen untereinander gekennzeichnet sein.</p> <p>Ein solches Klima ist u.a. dadurch charakterisiert, dass Lern- und Leistungssituationen deutlich getrennt werden. Insbesondere in vorangehenden Lernsituationen ist es wichtig, die SchülerInnenäußerungen ernst zu nehmen und mit ihnen produktiv und konstruktiv umzugehen. Auf diese Weise trägt das Aufgreifen der Beiträge dazu bei, dass die SchülerInnen eine gewisse Selbstwirksamkeit erfahren und erkennen, dass die Beteiligung am Unterrichtsgeschehen und der Austausch mit den MitschülerInnen förderlich für das eigene Lernen ist.</p> <p>Zu vermeiden sind sozial vergleichende Beurteilungen und Beschämungen. Förderlich für eine angstfreie Lernatmosphäre dürfte auch das Verhalten der Lehrperson wirken, wenn dieses von Freundlichkeit und Humor geprägt ist.</p> <p><i>Ankerbeispiel:</i></p> <p>„Was ich eben schon einmal gesagt habe: Das LehrerInnen-SchülerInnen-Gespräch ist suboptimal. Er stellt rhetorische Fragen, lässt die SchülerInnen überlegen, aber was die SchülerInnen sagen, wird direkt wieder abgewiesen. Er ist praktisch der Wissende, der weiß genau, was zu tun ist, was die SchülerInnen zu machen haben, ob die das wollen oder nicht. Dabei habe ich das Gefühl, dass er sich sogar teilweise über das Unwissen der SchülerInnen lustig macht, wenn er irgendwie lacht und dann sagt ‚das vergessen wir aber jetzt mal schnell wieder‘. Das finde ich nicht ermutigend.“</p>

## 3.2 Kategorien für fachliches (incl. fachdidaktisches) Wissen

Im vorangegangenen Teilkapitel 3.1 wurde erläutert, welche Facetten didaktisch-methodischen Wissens in dieser Untersuchung für die Bildung und Anwendung von Kategorien berücksichtigt werden. In diesem Teilkapitel sollen nun diejenigen Kategorien vorgestellt werden, die ich in meiner Qualitativen Inhaltsanalyse zur Identifizierung fachlichen (incl. fachdidaktischen) Wissens nutze (s. den Überblick über diese Kategorien in Tab. 10, S. 74). Um solche Kategorien herzuleiten (s. Kap. 3.2.3), knüpfe ich vorab exemplarisch an die gegenwärtige Diskussion über Notwendigkeit und Umfang dieser Wissensfacetten an (s. Kap. 3.2.1 und 3.2.2), wie sie im Bereich der Didaktik des Mathematikunterrichts und des naturwissenschaftlichen Unterrichts geführt wird. Abschließend werden die einzelnen Kategorien mit Definitionen, Beschreibungen und Ankerbeispielen vorgestellt (s. Kap. 3.2.4).

### 3.2.1 Das Vorgehen zur Bildung von Kategorien für fachliches und fachdidaktisches Wissen

Die in Kapitel 1.3 (s. S. 27ff.) skizzenhaft dargestellte Forschungs- und Diskussionslage hat exemplarisch die Schwierigkeiten deutlich gemacht, die im Rahmen der ausgewählten fachdidaktischen Bereiche (Mathematik, Physik) hinsichtlich der Begründung fachlicher und fachdidaktischer Wissensfacetten bestehen.

Für die im Folgenden zu bildenden Kategorien für fachliches und fachdidaktisches Wissen haben diese Schwierigkeiten zur Folge, dass ich nicht auf ein umfassendes und allseits akzeptiertes Konzept solchen Wissens zurückgreifen kann.

Wie kann man also mit derartigen Schwierigkeiten umgehen? – Ich versuche, mich einerseits an der oben umrissenen Forschungslage zu orientieren, da sie einen weiten Rahmen absteckt, in dem sich die folgenden Überlegungen zur Bildung von Kategorien für fachliches und fachdidaktisches Wissen bewegen (müssten). Über diesen allgemeinen Orientierungsrahmen hinaus erscheint es andererseits unumgänglich, die inhaltliche und methodische Faktur der zu analysierenden Unterrichtsstunde als Kriterium für die Wahl entsprechender Kategorien zu berücksichtigen, was bedeutet, dass letztlich nur solche fachlichen (incl. fachdidaktischen) Kategorien Anwendung finden, die im Hinblick auf die Analyse dieser spezifischen Unterrichtsstunde sinnvoll erscheinen.

Zusätzlich zum Stand der fachdidaktischen Diskussion um das Professionswissen und die spezifische Faktur der zu analysierenden Unterrichtsstunde erscheinen mir für die Bildung geeigneter Kategorien fünf weitere Aspekte relevant:

- 1) In Anlehnung an Plöger und Scholl (2014a) und Plöger et al. (2015) wird im Folgenden *keine* Differenzierung zwischen Fachwissen einerseits und fachdidaktischem Wissen andererseits vorgenommen. Beide Wissensbereiche werden als *ein* Kompetenzbereich *zusammengefasst* (s. Kap. 1, S. 22).<sup>71</sup>
- 2) Die zu bildenden Kategorien sollten eine Schnittmenge mit jenen Kategorien bilden, die sich in entsprechenden *Videostudien* bereits zur Erfassung der Analysekompetenz von Lehrpersonen als geeignet erwiesen haben. Da es sich bei dem zu analysierenden Unterricht um eine Physikstunde handelt, liegt es nahe, in dieser Hinsicht an die Videostudien der Arbeitsgruppe um Seidel und Prenzel anzuknüpfen (s. Seidel, 2003a; Seidel & Prenzel, 2004, 2006; Seidel, Prenzel, Rimmele, Schwindt et al., 2006; Seidel, Prenzel, Rimmele, Dalehefte et al., 2006).
- 3) Die Arbeiten der Forschungsgruppe um Seidel und Prenzel sind wiederum in einem größeren Kontext zu verorten. Dieser wird von der Frage der Bildungsqualität von Schule und Unterricht bestimmt. In diesem Zusammenhang spielen heutige *Bildungsstandards* eine zentrale Rolle. Sie sind in Form normativ-präskriptiver Texte formuliert und bestimmen den curricularen Rahmen schulischen Lehrens und Lernens. Wenn im Folgenden auf Texte dieser Art Bezug genommen wird, soll dadurch die „curriculare Validität“ (Döhrmann et al., 2010, S. 173, s. auch Riese et al., 2015, S. 59f., 68f.) bei der Bildung fachlicher (incl. fachdidaktischer) Kategorien gesichert werden. Als zentraler Orientierungsrahmen werden dabei die entsprechenden Beschlüsse der Kultusministerkonferenz (KMK) berücksichtigt: „Bildungsstandards im Fach Physik für den mittleren Schulabschluss“ (KMK 2004a), „Einheitliche Prüfungsanforderungen in der Abiturprüfung Physik“ (KMK 2004b) und „Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung“ (KMK 2013).
- 4) Infolge der unbefriedigenden Leistungen deutscher Schülerinnen und Schüler im internationalen Vergleich (s. z.B. *PISA 2000*, Prenzel et al., 2001; *TIMSS/III-Studie*, Baumert & Köller, 2000) kam die Diskussion um *Bildungsstandards* auf. Das schwache Abschneiden der SchülerInnen wird zu einem nicht geringen Teil auf die mangelnde Qualität von Unterricht zurückgeführt (s. z.B. Reusser & Pauli, 2003; Seidel & Prenzel, 2007). Als Beurteilungsmaßstab werden in der Regel Kriterien für Unterrichtsqualität (s. Kap. 1.2, S. 25) herangezogen, deren Wirkungen auf empirischen Weg im Wesentlichen nachgewiesen sind (s. z.B. im Überblick Helmke, 2012; Seidel & Shavelson, 2007). Diese Kriterien für Unterrichtsqualität stellen allerdings keine völlig neuen didaktisch-methodischen Forderungen dar, sondern sind gerade auch in der Tradition

---

71 Auf empirischem Wege wurde zwar ein hoher Zusammenhang zwischen fachdidaktischem Wissen und Fachwissen nachgewiesen, doch in der physikdidaktischen Forschung werden „trotz des Zusammenhangs zwischen fachdidaktischem Wissen und Fachwissen ... diese beiden Wissensfacetten [als] zwei theoretisch und empirisch trennbare Bereiche“ (Borowski, Kirschner, Liedtke, & Fischer, 2011, S. 4) aufgefasst (s. Riese & Reinhold, 2010). Ob und wie stark diese Trennung in der unterrichtlichen Praxis aussieht, ist bisher nicht eindeutig geklärt (s. Krauss et al., 2008). Vor allem im Hinblick auf Papier-Bleistift-Verfahren mag dieses Vorgehen nachvollziehbar erscheinen, je näher man aber an die unterrichtliche Praxis kommt (z.B. im Rahmen von Interviews), desto schwieriger wird es, die mit zunehmender Expertise voranschreitende Vernetzung dieser beiden Wissensbereiche sinnvoll zu trennen (s. z.B. Borowski et al., 2011; Krauss et al., 2008; Riese, 2009; Riese et al., 2015; Vogelsang & Reinhold, 2013). Des Weiteren ist es umstritten, ob fachdidaktisches Wissen überhaupt *inhaltsunabhängig* ist und somit getrennt vom Fachwissen erhoben werden kann (s. Gramzow et al., 2013, S. 13).

physikdidaktischen Denkens immer schon postuliert worden. Deshalb wird die Bildung entsprechender Kategorien im Folgenden auch durch Bezug auf die fachdidaktisch relevante Literatur argumentativ zu stützen sein, insbesondere durch Dokumente, die sich mit den Ausbildungsinhalten für das Lehramt Physik beschäftigen (s. z.B. Kircher, Girwitz, & Häußler, 2015; Mikelskis-Seifert & Rabe, 2007; Wiesner, Schecker, & Hopf, 2011).

- 5) Aussagen über zu erreichende LehrerInnen- und SchülerInnenkompetenzen, die in den Verlautbarungen der KMK (2004a, 2004b, 2013) und in den relevanten fachdidaktischen Publikationen zu finden sind, sind zum Teil eng an die Auffassungen angelehnt, die die Physik als wissenschaftliche Disziplin samt ihrer methodischen Vorgehensweise kennzeichnen. Deshalb erscheint es mir auch wichtig, die charakteristischen Merkmale physikalischer Forschung zu skizzieren, um daraus mögliche Kategorien für fachliches Wissen gewinnen zu können. Die entsprechenden Ausführungen dazu werden direkt im nächsten Abschnitt vorangestellt, bevor die Position der KMK referiert wird.

### 3.2.2 Gegenstand und Methode der Physik als indirekte Bezugsgröße der Kategorienbildung

#### 3.2.2.1 Physik als wissenschaftliche Disziplin

Da Physikunterricht immer auch ein Unterricht *über* Physik ist (s. Höttecke, 2007, 2008; Duit, 2015), erscheint zunächst ein Blick auf diese Wissenschaftsdisziplin selbst hilfreich, um erste Orientierungen für die Bildung möglicher bzw. notwendiger Kategorien zu bekommen. Ich beschränke mich dabei auf einige wenige charakteristische Merkmale, die diese Disziplin kennzeichnen: Physik als Erfahrungswissenschaft, Erfahrung von Naturphänomenen vs. physikalische Betrachtung von Natur, Mathematisierung physikalischer Größen, Erkenntnisfortschritt in der Physik als historischer Prozess.

#### Physik als Erfahrungswissenschaft

Physik ist eine empirische Wissenschaft, die sich mit Erscheinungen und Vorgängen der unbelebten Natur beschäftigt. Dabei erforscht sie die den Naturerscheinungen zugrunde liegenden Gesetzmäßigkeiten, deren Bausteine sowie ihre Wechselwirkungen untereinander.<sup>72</sup>

Als wissenschaftliche Disziplin im modernen Sinne existiert sie erst seit rund 400 Jahren: „Es war und bleibt im entscheidenden Maße das Verdienst Galileis, der Natur zu unterstellen, sie verhalte sich in *idealer* Weise so, dass bestimmte Größen von anderen funktional abhängen“ (Plöger, 2006, S. 22). Seit dieser Zeit lautet die übergeordnete Forschungsfrage der Physik: „Wie können die tatsächlichen Erscheinungen, die wir mit Hilfe unserer Sinnesorgane und mit Unterstützung von Instrumenten beobachten, auf einfache Begriffe zurückgeführt werden, die genauen Messungen zugänglich sind und bei der Formulierung quantitativer Gesetze gebraucht werden können?“ (Born, 1969, S. 12)

---

<sup>72</sup> Dieser eher traditionelle Ansatz, der vor allem auf den Erkenntnisgegenstand der Physik fokussiert, dient hier hauptsächlich zur Verdeutlichung der historischen Entwicklung der Disziplin Physik, die sich von den anderen klassischen Naturwissenschaften abzugrenzen versuchte. Diese Abgrenzung erscheint aber nicht mehr gänzlich zeitgemäß, da *interdisziplinäre* Verknüpfungen unberücksichtigt bleiben, wie z.B. die Biophysik und die Physikalische Chemie (s. Heering & Höttecke, 2011, S. 9; Duit, 1995b; Kircher, 2015a).

Um diese Forschungsprogrammatik einlösen zu können, benötigt die Physik eine entsprechende Terminologie. Aus begriffsanalytischer Sicht muss dabei zwischen sogenannten *Grundgrößen* (auch „Basisgrößen“) und *abgeleiteten Größen* unterschieden werden. Gesetzte Größen, die nicht auf andere Größen definitorisch zurückgeführt werden können, wie z.B. der Weg oder die Zeit, werden Grundgrößen genannt.<sup>73</sup> Aus diesen Grundgrößen lassen sich weitere Größen bilden, wie z.B. die Geschwindigkeit, die aus den Grundgrößen Weg (Länge) und Zeit abgeleitet wird.

Damit Grundgrößen und abgeleitete Größen hinreichend genau bestimmt werden können, werden entsprechend präzise *Messverfahren* benötigt, wobei die Messung selbst auf einem Vergleich mit einer festgelegten Einheit beruht.

### **Erfahrung von Naturphänomenen vs. physikalische Betrachtung von Natur**

Physikalisches Denken nimmt seinen *Ausgangspunkt* von natürlichen Phänomenen (z.B. Mond- und Sonnenfinsternis, Kraftwirkungen usw.), die dem Laien aus dem Alltagserleben heraus mehr oder weniger vertraut sind, aber auch zu dem führen können, was Wagenschein als „ehrfürchtiges Staunen“ (Wagenschein, 1965, S. 85) bezeichnet hat. Diese alltägliche Erfahrung von Naturphänomenen geht nicht selten einher mit emotionalen Erlebnissen (z.B. „wohltuende“ Wärme, die Faszination beim Betrachten eines Sonnenauf- oder -untergangs) und ist oft durchzogen von anthropomorphisierenden Deutungen (Fensterscheiben „schwitzen“, Unterdruck als „Saugen“).

Diese subjektiv gefärbte Sicht auf Naturphänomene wird in den Naturwissenschaften aber ausgeblendet. Nach Dilthey konstituiert die Naturwissenschaft die „Natur“ als ihren Gegenstand dadurch, dass „das lebendige Gefühl, in dem wir sie genießen, immer mehr zurücktritt hinter das abstrakte Auffassen derselben nach den Relationen von Raum, Zeit, Masse, Bewegung. Alle diese Momente wirken dahin zusammen, daß der Mensch sich selbst ausschaltet, um aus seinen Eindrücken diesen großen Gegenstand Natur als eine Ordnung nach Gesetzen zu konstruieren“ (Dilthey, 1979, S. 83).

### **Das Experiment als zentrale Methode physikalischer Forschung**

Im Vergleich zur alltäglichen Erfahrung von Naturphänomenen zeichnet sich die physikalische Sicht also durch einen anderen, „selektiven, aspekthaften Zugriff auf Natur“ (Plöger, 2006, S. 23) aus. Phänomene werden dadurch als „Tatsachen“ im Lichte theoretischer Auffassungen gedeutet. Diese theoretischen Vorstellungen (begrifflich in Grundgrößen und abgeleitete Größen gefasst) sind rein hypothetischer Natur. Aus ihnen werden auf deduktivem Wege Voraussagen abgeleitet, über deren Angemessenheit das *Experiment* empirisch entscheidet. Experimentieren stellt somit die zentrale Methode physikalischer Forschung dar. Dabei wird in einem allgemeinen Sinne unter *Experimentieren* das aktive „Eingreifen und Gestalten von Natur und Technik unter kontrollierten Rahmenbedingungen“ (Heering & Höttecke, 2011, S. 13) verstanden, um so funktionale Zusammenhänge zwischen spezifischen Variablen untersuchen zu können. Zu diesem Zweck werden physikalische Größen in eine mathematisierte Form überführt, wobei insbesondere die von Newton entwickelte Infinitesimalrechnung herangezogen wird.

73 Nach Kircher (2015b) benötigt die „klassische Physik“ lediglich sieben Grundgrößen, aus denen sich sämtliche anderen physikalischen Größen ableiten lassen (s. Kircher, 2015b, S. 34f.). Diese Grundgrößen sind: Länge (Meter), Masse (Kilogramm), Zeit (Sekunde), Elektrische Stromstärke (Ampere), Temperatur (Kelvin), Stoffmenge (Mol) und Lichtstärke (Candela).

### Erkenntnisfortschritt in der Physik als historischer Prozess

Die theoriegeleitete und experimentelle Vorgehensweise im Bereich Physik zieht Fragen des Erkenntnisfortschrittes nach sich. Durch experimentelle Überprüfung werden Theorien ständig verifiziert, modifiziert, aber auch falsifiziert und dementsprechend eliminiert. Die über derartige Prozesse generierten neueren Theorien können aber keinen Anspruch auf Wahrheit geltend machen. Ihr Erklärungs- und Prognosegehalt übersteigt zwar den der älteren Theorien, jedoch nur solange, bis der durch das Vorliegen zahlreicher und gravierender Anomalien erzeugte Druck berechtigten Zweifel an der betreffenden Theorie aufkommen lässt (s. Kircher, 2015d).

In den letzten Jahrzehnten sind viele Arbeiten erschienen, die den *konstruktiven* und damit geschichtlich bedingten Charakter naturwissenschaftlicher Theoriebildung betonen (s. z.B. Kuhn, 1979; Sneed, 1971; Stegmüller, 1979), wobei vor allem die Arbeiten von Thomas Kuhn große Beachtung erzielt haben. Aus Kuhns Sicht ist empirische Forschung ein überaus pragmatisches Vorgehen, das teilweise auf einem irrationalen Fundament beruhe, denn die ForscherInnen müssten sich zu einem (theoretisch und methodologisch bestimmten) Paradigma *bekennen*, um sich dadurch als Mitglied der scientific community zu identifizieren.

Kommt es zu einer „wissenschaftlichen Revolution“, durch die ein bestehendes Paradigma von einem neuen, mit der alten Theorie inkommensurablen, Paradigma abgelöst wird, führt dies nicht zwangsläufig dazu, dass alle Wissenschaftler vom neuen Paradigma überzeugt sind, obwohl ihnen die „kritischen“ empirischen Belege bestens bekannt sind. Im Hinblick auf die Etablierung von neuem Wissen stellt Max Planck deshalb am Ende seiner wissenschaftlichen Laufbahn mit Ernüchterung fest: „Eine neue wissenschaftliche Wahrheit pflegt sich nicht in der Weise durchzusetzen, daß ihre Gegner überzeugt werden und sich als belehrt erklären, sondern vielmehr dadurch, daß die Gegner allmählich aussterben und daß die heranwachsende Generation von vornherein mit der Wahrheit vertraut gemacht ist“ (Planck, 1948, S. 22).

Die Arbeiten Kuhns, in denen er das fortwährende Wechselspiel von Theorie und Experiment betont und dadurch Fragen des Erkenntnisfortschritts beleuchtet, machen deutlich, dass physikalisches Wissen nicht die natürliche Wirklichkeit abbildet bzw. abbilden kann. Vielmehr werden Begriffe auf „Natur“ übertragen, um sie dadurch in Form physikalischer Konstrukte beschreiben zu können. Solche Konstrukte haben aber *artificialen* Charakter; sie sind geschichtlich bedingt und somit kontingent. Dementsprechend stellt naturwissenschaftliches Wissen eine „tentative Konstruktion“ (Duit, 1995a, S. 912) dar, die impliziert, „dass das heutige naturwissenschaftliche Wissen eine Momentaufnahme darstellt und künftig (eventuell) anders konstruiert sein“ (Plöger, 2006, S. 25) könnte. Oder wie es Kircher (2015b) formuliert: „Die ‚Wahrheit der Physik‘ ist ein unendlicher, schwieriger Weg, der nur zu vorläufigen, nicht zu endgültigen Resultaten führt“ (Kircher, 2015b, S. 32).<sup>74</sup>

Dieser kurze Exkurs zum Selbstverständnis und zur grundsätzlichen Arbeitsweise der Disziplin Physik verweist auf Aspekte, die auch das Lehren von Physik im Kontext von Schule bestimmen. Konstitutiv für jeglichen Physikunterricht ist der Erwerb und Umgang

---

<sup>74</sup> Derzeit gibt es eine rege Debatte darüber, inwieweit solche erkenntnis- und wissenschaftstheoretischen Fragen, wie sie hier in knapper Form angesprochen wurden, fruchtbar in den Unterricht integriert werden können (s. z.B. Kircher, 2015b).

mit Begriffen, die aus der zugehörigen wissenschaftlichen Disziplin stammen. Lehrpersonen, die dieses Fach unterrichten, müssten daher über entsprechendes begriffliches Wissen in präziser Weise verfügen. Zu einem tiefen Verständnis physikalischer Begriffe sollten sie auch deren historische Genese – zumindest auf exemplarische Weise – kennen. Darüber hinaus müssen sie mit der Methode des Experimentierens vertraut sein, die für den Erkenntnisprozess von SchülerInnen einen zentralen Stellenwert im Unterricht einnimmt. Diese Anforderungen an Lehrpersonen beziehen sich nicht nur auf die Planung und Durchführung von Unterricht, sondern sind auch unabdingbar für ein angemessenes Sprechen und Reflektieren über Unterricht, wie es bei dessen Analyse der Fall ist. Deshalb werde ich auf diese gerade beschriebenen Aspekte im Folgenden bei der Bildung von Kategorien für die Analyse der ausgewählten Unterrichtsstunde zurückgreifen.

### 3.2.2.2 Physik als Unterrichtsfach: Schüler- und Lehrerkompetenzen

Das im Bereich der Schule unterrichtete Fach Physik ist kein direktes Abbild der im Vorangegangenen in pointierter Form präsentierten Disziplin Physik – sozusagen eine „Physik im Kleinen“. Zwar ist das grundlegende Selbstverständnis dieses Unterrichtsfaches auch an der universitären Disziplin orientiert, aber es unterliegt in erster Linie einem *Allgemeinbildungsauftrag*, den jedes Fach auf seine spezifische Weise unter Maßgabe fachlicher Leitziele erfüllen soll (s. dazu KMK, 2004a; Plöger, 2006).

Fachunterricht soll SchülerInnen unterstützen, ihr Selbst- und Weltverständnis zu verbessern, um dadurch ihr gegenwärtiges und vor allem zukünftiges Leben in Selbstbestimmung und sozialer Verantwortung führen zu können. Daher müssen sich die SchülerInnen ein breites und tiefes Wissen aneignen, das sowohl für die persönliche Lebensführung als auch für die Partizipation an gesellschaftlichen Prozessen unerlässlich ist.

Diese Positionierung des Physikunterrichts vor dem Hintergrund der Fachdisziplin einerseits und dem Bildungsauftrag der allgemeinbildenden Schule andererseits prägt dann auch die grundsätzliche Ausrichtung der von der Kultusministerkonferenz beschlossenen *Bildungsstandards* für den *mittleren Schulabschluss*: Im Physikunterricht wird „eine Grundlage für die Auseinandersetzung der jungen Menschen mit naturwissenschaftlichen Themen und ihren gesellschaftlichen Zusammenhängen gelegt. Zudem leistet er einen Beitrag zu anderen Fächern und zur Vorbereitung auf technische Berufe bzw. weiterführende Bildungsgänge und ermöglicht damit ein anschlussfähiges Orientierungswissen“ (KMK, 2004a, S. 6).

Die angestrebten Ziele naturwissenschaftlicher Grundbildung bestehen darin, „Phänomene erfahrbar zu machen, die Sprache und Historie der Naturwissenschaften zu verstehen, ihre Ergebnisse zu kommunizieren sowie sich mit ihren spezifischen Methoden der Erkenntnisgewinnung und deren Grenzen auseinander zu setzen. Dazu gehört das theorie- und hypothesengeleitete naturwissenschaftliche Arbeiten, das eine analytische und rationale Betrachtung der Welt ermöglicht“ (ebd.).

Eine nähere Bestimmung erfahren diese Ziele durch die Angabe von SchülerInnenkompetenzen in Form von vier Kompetenzbereichen, mit denen „die notwendige physikalische Grundbildung“ (KMK, 2004a, S. 7) beschrieben wird:



**Tab. 23:** Kompetenzbereiche im Fach Physik (nach KMK, 2004a)

<b>Kompetenzbereiche im Fach Physik</b>	
Fachwissen	Physikalische Phänomene, Begriffe, Prinzipien, Fakten, Gesetzmäßigkeiten kennen und Basiskonzepten zuordnen
Erkenntnisgewinnung	Experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen
Kommunikation	Informationen sach- und fachbezogen erschließen und austauschen
Bewertung	Physikalische Sachverhalte in verschiedenen Kontexten erkennen und bewerten

Diese Kompetenzbereiche werden schließlich durch Angabe von zu erreichenden SchülerInnenleistungen in Form von „Regelstandards“ (ebd., S. 11) konkretisiert.

Für den *Bereich der Sekundarstufe II* hat die Kultusministerkonferenz „Einheitliche Prüfungsanforderungen in der Abiturprüfung Physik“ (KMK, 2004b) beschlossen, mit denen an die Bildungsintentionen der Sekundarstufe I (s. oben) angeknüpft werden soll. Dabei wird dann auch ein stärkerer Bezug zur universitären Disziplin gefordert:

Die Bedeutung des Unterrichtsfaches Physik erschließt sich aus den Merkmalen der Fachwissenschaft Physik. Physik

- ist eine theoriegeleitete Erfahrungswissenschaft
- betrachtet die Natur unter bestimmten Aspekten (Aspektcharakter)
- hat einen hohen Grad an Formalisierung und Mathematisierung
- entwickelt ein spezifisches Methodenrepertoire
- hat starke Anwendungsbezüge und hohe gesellschaftliche Relevanz
- ist ein historisch-dynamischer Prozess.

(KMK, 2004b, S. 2)

Die in der Abiturprüfung von SchülerInnen nachzuweisenden Kompetenzen werden durch vier Bereiche definiert, die als Vertiefung der bereits oben für die Sekundarstufe I genannten Kompetenzen zu verstehen sind:

- Kompetenzbereich Fachkenntnisse: Physikalisches Basiswissen (z.B. Begriffe, Größen, Gesetze) zu den zentralen physikalischen Teilgebieten
- Kompetenzbereich Fachmethoden: Erkenntnismethoden der Physik sowie Fachmethoden beschreiben und nutzen
- Kompetenzbereich Kommunikation: In Physik und über Physik kommunizieren
- Kompetenzbereich Reflexion: Über die Bezüge der Physik reflektieren.

(KMK, 2004b, S. 3f.)

Während sich die angeführten Bildungsinhalte auf die zu erwerbenden Kompetenzen und Kompetenzbereiche der SchülerInnen beziehen, werden in Entsprechung dazu auch die Kompetenzprofile aufgezählt, die Lehrpersonen im Rahmen der universitären Ausbildung erwerben sollen, damit sie befähigt sind, ihre SchülerInnen bei der Aneignung der oben beschriebenen Kompetenzen effektiv zu unterstützen.

Diese fachbezogenen Kompetenzen von LehrerInnen wurden von der Kultusministerkonferenz durch „Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung“ (KMK, 2013) in Anlehnung an die Forderungen

der Gesellschaft für Fachdidaktik (GFD, 2004) sowie der Deutschen Physikalischen Gesellschaft (DPG, 2006) formuliert:

Die Studienabsolventinnen und -absolventen verfügen über die grundlegenden Fähigkeiten für gezielte und nach wissenschaftlichen Erkenntnissen gestaltete Vermittlungs-, Lern- und Bildungsprozesse im Fach Physik. Sie

- verfügen über anschlussfähiges physikalisches Fachwissen, das es ihnen ermöglicht, Unterrichtskonzepte und -medien fachlich zu gestalten, inhaltlich zu bewerten, neuere physikalische Forschung in Übersichtsdarstellungen zu verfolgen und neue Themen in den Unterricht einzubringen,
- sind vertraut mit den Arbeits- und Erkenntnismethoden der Physik und verfügen über Kenntnisse und Fertigkeiten im Experimentieren und im Handhaben von (schultypischen) Geräten,
- kennen die Ideengeschichte ausgewählter physikalischer Theorien und Begriffe sowie den Prozess der Gewinnung physikalischer Erkenntnisse (Wissen über Physik) und können die gesellschaftliche Bedeutung der Physik begründen,
- verfügen über anschlussfähiges fachdidaktisches Wissen, insbesondere solide Kenntnisse fachdidaktischer Konzeptionen, der Ergebnisse physikbezogener Lehr-Lern-Forschung, typischer Lernschwierigkeiten und Schülervorstellungen in den Themengebieten des Physikunterrichts, sowie von Möglichkeiten, Schülerinnen und Schüler für das Lernen von Physik zu motivieren,
- kennen Möglichkeiten zur Gestaltung von Lernarrangements unter dem besonderen Gesichtspunkt heterogener Lernvoraussetzungen und kennen den Stand physikdidaktischer Forschung und Entwicklung zum fachbezogenen Lehren und Lernen in inklusiven Lerngruppen,
- verfügen über erste reflektierte Erfahrungen im Planen und Gestalten strukturierter Lehrgänge (Unterrichtseinheiten) sowie im Durchführen von Unterrichtsstunden.

(KMK, 2013, S. 50)

Die in dieser Auflistung benannten Kompetenzfacetten decken sich z.T. mit den oben (s. Kap. 3.2.2.1) bereits im Zusammenhang mit den charakteristischen Merkmalen der Disziplin Physik angesprochenen Kompetenzen von Lehrpersonen: tiefes Verständnis physikalischer Begriffe, Beherrschen der experimentellen Methode und Kenntnisse über die historische Genese physikalischer Erkenntnisse. In den gerade beschriebenen curricularen Anforderungen werden zudem weitere Arbeitsmethoden der Physik genannt, zu denen z.B. die grafische Darstellung physikalischer Sachverhalte oder die Gestaltung von Übersichtsdarstellungen gehören. Darüber hinaus kommen fachdidaktische Kompetenzfacetten ins Spiel (Kenntnis fachdidaktischer Konzeptionen, Kenntnisse physikbezogener Lehr-Lern-Forschung und der professionelle Umgang mit Lernschwierigkeiten und SchülerInnenvorstellungen über physikalische Sachverhalte).

### 3.2.3 Auswahl der Kategorien für fachliches und fachdidaktisches Wissen

Nach diesen knappen Ausführungen zur wissenschaftlichen Disziplin Physik und den curricularen Anforderungen an das daran orientierte Unterrichtsfach stellt sich nun die Frage, wie das im Vorangegangenen abgesteckte weite Feld möglicher Wissensfacetten auf solche Kategorien für fachliches (und fachdidaktisches) Wissen reduziert werden kann, die für die Analyse der ausgewählten Unterrichtsstunde aussagekräftig und zugleich tauglich zur Prüfung des Konstrukts Analysekompetenz sind.

Mit Blick auf die zu analysierende Unterrichtsstunde lassen sich *sieben* Kategorien (= Kompetenzfacetten) bestimmen, die deduktiv „gesichert“ sind, d.h. die allesamt kompatibel mit den zuvor beschriebenen Merkmalen der Disziplin Physik, mit den curricular

eingeforderten SchülerInnenkompetenzen und insbesondere mit den im Hochschulstudium zu erwerbenden LehrerInnenkompetenzen sind und als *Minimalkonsens* in den theoretischen wie empirischen Arbeiten im Bereich der Physikdidaktik (s. die Belege unten) akzeptiert sind. Das sind die Kategorien:

- Physikalische Begriffsbildung
- Durcharbeiten von Fachbegriffen
- Experimentieren als zentrale Fachmethode
- (weitere) fachliche Arbeitsweisen: Phänomene beobachten und genau analysieren; Messdaten grafisch darstellen; kritisches Beurteilen von Messwerten; qualitative Zusammenhänge (ansatzweise) mathematisieren
- Historische Genese physikalischen Wissens
- Produktiver Umgang mit Fehlern
- Lernen als konstruktive Tätigkeit von SchülerInnen

Im weiteren Verlauf werden diese Kategorien definiert, beschrieben und an Ankerbeispielen festgemacht (s. Kap. 3.2.4). Vorangestellt seien aber (zusätzlich zu den in Kap. 3.2.2.1 und Kap. 3.2.2.2 angeführten Argumenten für die jeweiligen Kategorien) einige Hinweise zur fachlichen und fachdidaktischen Legitimation dieser Kategorien.

### **Physikalische Begriffsbildung**

Die klare und verständliche Einführung von Fachbegriffen zählt zu den Standardsituationen des Physikunterrichts, denn in Verbindung mit und in Ergänzung zur Alltagssprache (s. dazu Fluck, 1996; Kattmann, 2006; Rincke, 2012; Rincke & Leisen, 2015) ermöglicht erst die Verfügbarkeit solcher Begriffe die angemessene Kommunikation über Sachverhalte unter physikalischen Aspekten (s. KMK, 2004a, S. 12). Das von den SchülerInnen anzueignende Begriffsrepertoire bezieht sich vor allem auf Phänomene, Fakten, physikalische Grundgrößen und abgeleitete Größen, aber auch auf funktionale, gesetzmäßige Zusammenhänge.

Damit die SchülerInnen ein tiefes Verständnis der physikalischen Sachverhalte erwerben können, müssen die gelernten Begriffe hinreichend vernetzt sein. Erst durch diese Vernetzung ist *kumulatives* Lernen möglich. Weber (2002) unterscheidet dabei zwischen interner und externer vertikaler Vernetzung. Mit *interner vertikaler* Vernetzung ist die Verknüpfung von (Teil-)Begriffen *innerhalb* eines Inhaltsbereiches wie Mechanik oder Optik gemeint. Dagegen stellt die *externe vertikale* Vernetzung Beziehungen *zwischen* solchen Bereichen dar und wird vorwiegend durch sogenannte „Basiskonzepte“ (z.B. Wechselwirkung, Erhaltung, Energieumwandlung) angestrebt.

Lehrpersonen müssen diese physikalischen Begriffe nicht nur verstanden und verinnerlicht haben, sondern auch noch darüber hinaus über hinreichendes Wissen bezüglich der entsprechenden SchülerInnenvorstellungen verfügen (Müller, Wodzinski, & Hopf, 2004; Wiesner, 2008). Sollen SchülerInnen nun etwas Neues (hier physikalische Sachverhalte) lernen, dann können sie zunächst nur an ihr relevantes Alltagswissen anknüpfen (s. Duit, 2015). Daraus können Probleme bzw. Lernschwierigkeiten resultieren, denn die zu erarbeitenden physikalischen Begriffe stellen, im Gegensatz zu den emotional wie anthropo-

morphisierend gefärbten Begriffen der SchülerInnen, eine durch Messverfahren objektivierte Erfassung der Natur dar.<sup>75</sup> Dieser Unterschied zwischen natürlicher, sinnlich affizierter Erfahrung und objektivierter Erfassung erfordert von den SchülerInnen letztlich einen *Konzeptwechsel* (conceptual change), der ihnen mehr oder weniger große Schwierigkeiten bereitet und selten „restlos“ erfolgt (s. Chinn & Brewer, 1998; Duit, 1993, 1994, 2006, 2015; Duit & Treagust, 2003; Hewson, Beeth, & Thorley, 1998; Jung, 1986).<sup>76</sup>

Das mit einem Konzeptwechsel verbundene Ziel besteht darin, „den Schülerinnen und Schülern verständlich zu machen, daß in bestimmten Situationen die Alltagsvorstellungen keine ausreichende Orientierung mehr bieten und daß in diesen Situationen die naturwissenschaftlichen Vorstellungen herangezogen werden müssen“ (Duit, 1996, S. 146). Prinzipiell lassen sich zwei Arten von Lernwegen unterscheiden, um „von bestimmten bereits vorhandenen Vorstellungen zu fachlichen Vorstellungen zu gehen“ (Duit, 1995a, S. 913): Zum einen kann ein *kontinuierlicher* Lernweg gewählt werden. Dabei „wird nach Möglichkeiten gesucht, mit Erweiterungen und kleineren Revisionen der vorhandenen Vorstellungen auszukommen“ (ebd.). Zum anderen kann aber auch ein *diskontinuierlicher* Lernweg geboten sein, wenn die Vorstellungen der SchülerInnen „einer grundlegenden Revision“ (ebd.) bedürfen. Hierbei bilden die bestehenden, teilweise recht deutlich von der physikalischen Perspektive abweichenden SchülerInnenvorstellungen den Ausgangspunkt für den angestrebten Konzeptwechsel, der durch *kognitive Konflikte* bei den SchülerInnen initiiert werden soll.

Dieser Konzeptwechsel kann in methodischer Hinsicht durch eine entsprechende Sequenzierung des Unterrichts unterstützt werden: Ausgang vom Phänomen, Thematisierung der SchülerInnenvorstellungen, Einführung des physikalischen Begriffs, Anwendung (s. Duit, 1993; Duit, Treagust, & Widodo, 2008). Innerhalb dieser Sequenzierung spielt auch der Einsatz von *Modellen* eine wichtige Rolle als Lernhilfe (Fischler, 1997; Gilbert, 2004; Gilbert & Boulter, 1998; Groslight, Unger, & Jay, 1991; Harrison & Treagust, 1996; Kircher, 2015c; Meisert, 2008; Mikelskis-Seifert & Fischler, 2003), wobei auf SchülerInnen-seite oft eine Vermischung von Realitäts- und Modellebene zu verzeichnen ist. Daher sollten Lehrpersonen die Analogie- und Lernmöglichkeiten von Modellen kennen (Leisner, 2005; Mikelskis-Seifert & Leisner, 2003), um deren erkenntnistheoretische Grenzen wissen (Kircher, 1995, 2015c) und diese im Unterricht – adaptiert an das Verständnis der SchülerInnen – auch explizit thematisieren. Diese Forderung ergibt sich insbesondere aus Untersuchungen, die zeigten, dass bei Lehrpersonen entsprechende Defizite im Verständnis und im Umgang mit Modellen sowie hinsichtlich ihrer erkenntnistheoretischen Relativierung vorliegen (s. z.B. Justi & Gilbert, 2003).

<sup>75</sup> Diese Alltagsvorstellungen der SchülerInnen beziehen sich nicht nur auf zu erlernende physikalische Phänomene, Begriffe und Prinzipien. Nach Duit (2015) können Alltagsvorstellungen zusätzlich noch die Vorstellungen über das „Wesen“ der Physik sowie „über das (eigene) Lernen“ umfassen. Demnach sind SchülerInnen eher „naive Realisten“, die von einer identischen Abbildung der Wirklichkeit durch die Physik ausgehen und darüber hinaus unter Lernen die „schlichte Übernahme und Speicherung von Wissen“ (Duit, 2015, S. 661) verstehen. In der Kategorie „Lernen als konstruktive Tätigkeit des Schülers“ wird sich diesen beiden Aspekten noch einmal ausführlicher gewidmet.

<sup>76</sup> Posner, Strike, Hewson und Gertzog (1982) postulieren vier Bedingungen, die notwendig seien, um einen Konzeptwechsel („conceptual change“) anzuregen: *dissatisfaction* (Unzufriedenheit mit vorhandenen Vorstellungen); *intelligible* (Verständlichkeit der neuen Vorstellung); *plausible* (neue Vorstellung muss plausibel sein); *fruitful* (fruchtbare bzw. erfolgreiche Anwendung der neuen Vorstellung).

### Durcharbeiten von Fachbegriffen

Damit *träges Wissen* (s. z.B. Gruber & Renkl, 2000; Gruber et al., 2000; Renkl, 1996, 1998; Renkl & Nückles, 2006) vermieden und der Aufbau *anwendbaren* Wissens gefördert wird, sollten physikalische Begriffe *alltagsnah* eingeführt werden (Girwidz & Berger, 2008; Labudde, 1997). Durch die Einbettung in „sinnstiftende Kontexte“ (Muckenfuß, 1995, 2004) wird nicht nur das Interesse von SchülerInnen stärker geweckt, sondern auch die spätere Anwendung der betreffenden Begriffe erleichtert (Mandl & Kopp, 2005).

Die bei der Einführung wie bei der Anwendung von Begriffen verwendeten Kontexte dürfen aber keine isolierten Größen bilden, sondern müssen mit den Merkmalen des jeweiligen physikalischen Begriffs in Zusammenhang gebracht werden. Im Sinne exemplarischen Lehrens und Lernens müssen Lehrpersonen über die Kompetenz verfügen, Kontexte in Form prägnanter Beispiele zu präsentieren und dadurch Strukturisomorphie zwischen Beispiel und entsprechendem Begriff herzustellen. Empirische Befunden deuten aber darauf hin, dass Lehrpersonen diesen Sachverhalt zu selten berücksichtigen (s. Bennet, Hogarth, & Lubben, 2003). In den letzten Jahren konnte die lernunterstützende Wirkung kontextbezogenen Arbeitens durch die fachdidaktische Forschung nachgewiesen werden (s. für die Fächer Biologie, Chemie und Physik z.B. Bayrhuber et al., 2007; Gräsel, Nentwig, & Parchmann, 2005; Mikelskis-Seifert & Duit, 2010).

Im Anschluss an die kontextbezogene Einführung physikalischer Begriffe sollte aber nicht unmittelbar die Anwendung folgen, da die erlernten Begriffe und deren Merkmale noch zu sehr mit dem eingangs gewählten beispielhaften Kontext verbunden sind. Daher ist darauf zu achten, dass die einzelnen Merkmale und ihre spezifische Beziehung zueinander *klarer* hervorgehoben werden. Ohle, Fischer und Kauertz (2011) markieren diesen entscheidenden Schritt als dritten von insgesamt fünf Schritten: „Aktivierung des Vorwissens, Einführung des neuen Konzepts, Beschreiben der wesentlichen Merkmale des neuen Konzepts, Anwendung des Konzepts, Verknüpfung mit anderen Konzepten, Übertragung in andere Kontexte“ (ebd., S. 364).

In der Lehr-Lerntheorie Aebli (1993a) wird die Funktion dieses Schrittes, die Phase des *Durcharbeitens*, noch deutlicher betont. Nach Aebli sind die zu lernenden Begriffe nach ihrer erstmaligen Erarbeitung noch relativ instabil (s. Aebli, 1993a, S. 269). SchülerInnen können demnach noch nicht hinreichend die konkreten Merkmale des Kontextes, in dem die Begriffe eingeführt wurden, von den wesentlichen, abstrakten Merkmalen der Begriffe selbst unterscheiden, sodass auch ein (möglicher) Transfer des betreffenden Begriffes auf einen anderen Kontext nicht oder nur ansatzweise gelingt. Damit die Beziehungen zwischen den Begriffselementen deutlicher hervortreten können, sollten Lehrpersonen die Phase des Durcharbeitens eines Begriffes zwischen Erarbeitung und Transfer stärker berücksichtigen, um auf diese Weise die Basis für die Dekontextualisierung der jeweiligen begrifflichen Struktur zu legen. Insofern zielt das im Durcharbeiten auf eine Beweglichkeit im Denken und Handeln, die die entscheidende Voraussetzung für den Transfer des Begriffes auf weitere Situationen darstellt (s. Baumert & Köller, 2000, S. 274).

### Experimentieren als zentrale Fachmethode

Das Experiment nimmt bei der Gewinnung physikalischer Erkenntnisse eine zentrale Stellung ein. Es stellt die wesentliche Erkenntnisquelle der Physik dar und wird hier deshalb im Vergleich zu weiteren typischen Arbeitsweisen (s. dazu den nächsten Abschnitt) gesondert herausgehoben.

Aufgrund dieser exponierten Stellung wird das Experimentieren in der Fachdidaktik Physik seit Jahrzehnten als wichtiger unterrichtsmethodischer Baustein gesehen (s. z.B. Duit, 1995a; Hammann, Phan, Ehmer, & Bayrhuber, 2006; Girwidz, 2015; Magnusson, Krajcik, & Borko, 1999; Mikelskis-Seifert & Leisner-Bodenthin, 2007; Riese & Reinhold, 2009; Tepner et al., 2012; Tesch, 2005; Wilke, 1993a, 1993b).<sup>77</sup>

Die mit dem Einsatz des Experiments im Unterricht verbundenen Erwartungen wurden und werden jedoch nur selten eingelöst (Duit, 1993). In einem Überblicksartikel berichtet Harlen (1999), dass SchülerInnenexperimente als Methode der Erkenntnisgewinnung nicht effektiv zum Einsatz kommen, da SchülerInnen zu wenig zur eigenständigen Konstruktion von Experimenten aufgefordert werden. Dies hätte dann zur Folge, dass die Ausführung eher rezeptartigen Anweisungen folge und inhaltlich wenig anspruchsvoll sei.

Diese Feststellung wird für die Bundesrepublik Deutschland im Wesentlichen nicht nur durch die Ergebnisse der *TIMS/III-Studie* bestätigt (s. Baumert & Köller, 2000), sondern auch mittels neuerer Untersuchungen in Form von Videoanalysen. Tesch und Duit (2004) sowie die ForscherInnengruppe um Seidel (s. z.B. Seidel, Prenzel, Rimmele, Schwindt et al., 2006) konnten nachweisen, dass SchülerInnen nur selten die Möglichkeit zur eigenständigen Planung und Auswertung von Experimenten bekommen (s. auch Duit, 2006, S. 18). Es zeigte sich, „dass rund zwei Drittel der Unterrichtszeit (64%) in, teilweise lockerem, Zusammenhang mit dem Experimentieren gestaltet werden ... Bei einzelnen Lehrpersonen sind es sogar über 80% der Unterrichtszeit. Es wird insgesamt mehr Zeit für Vor- und Nachbereiten als für das eigentliche Experimentieren aufgewendet.“ (Tesch & Duit, 2004, S. 59) Gleichzeitig konnten die Autoren nachweisen, dass die *Gesamtdauer* der Experimentierphasen (Experimentieren einschließlich Hinführung, Planung und Auswertung) in einem deutlichen Zusammenhang zu den erzielten SchülerInnenleistungen steht: „Je stärker der Unterricht durch Experimente bestimmt ist, desto besser ist die Leistungsentwicklung“ (ebd., S. 64). Keine Bedeutung für die Leistungen haben dagegen die zeitliche Länge der Experimentierphase und auch nicht die Zahl durchgeführter Experimente.

Diese Ergebnisse lenken den Blick auf eine wichtige Kompetenz von Lehrpersonen: Sie müssen den Unterricht so gestalten, dass die praktische Ausführung von Experimenten in einen übergreifenden *Prozess* integriert ist (Baumert & Köller, 2000; Girwidz, 2015; Harlen, 1999; Hofstein & Lunetta, 1982, 2004; Lunetta, 1998; Fischer et al., 2003; Tesch & Duit, 2004), in dem die Phasen *Forschenden Lernens* (Bell, 2007; Fries & Rosenberger, 1973; Plöger, 1983) ineinander verzahnt sind: Problemstellung, Hypothesenbildung, Versuchsplanung, Versuchsdurchführung und -auswertung.

---

<sup>77</sup> Eine systematische Beschreibung des Experiments und des Experimentierens im Physikunterricht nimmt Girwidz (2015) vor. Dabei unterscheidet er einerseits zwischen den *funktionalen Aspekten*, die das Einsatzspektrum und die entsprechende Zielsetzung von Experimenten im Unterricht beschreiben und andererseits zwischen verschiedenen *Formen* von Experimenten, mit denen „unterschiedliche methodische Möglichkeiten und/oder Anforderungsprofile verknüpft sind“ (Girwidz, 2015, S. 232f.).



### (weitere) fachliche Arbeitsweisen

Im Rahmen des Forschenden Lernens werden weitere Kompetenzen benötigt, die für ein verständnisorientiertes Experimentieren unabdingbar sind. Hierzu gehören insbesondere: Phänomene beobachten und genau analysieren, Messdaten grafisch darstellen; kritisches Beurteilen von Messwerten; qualitative Zusammenhänge (ansatzweise) mathematisieren (s. Duit, 2003; Duit, Gropengießer, & Stäudel, 2004; Häußler, 1973; Klopfer, 1971; KMK, 2004a; Stadler, 2009).

#### – Phänomene beobachten und genau analysieren

Es ist ein Verdienst Martin Wagenscheins (1965, 1968, 1975), dass dem Phänomen im Physikunterricht eine zentrale Stellung beigemessen wird. In den von Wagenschein empfohlenen „Regeln“ zum exemplarischen Lehren und Lernen heißt es beispielsweise:

- Erst das Naturphänomen, dann das Laborphänomen;
  - Erst ‚qualitativ‘, dann quantitativ;
  - Erst das Phänomen, dann die Theorie und die Modellvorstellung;
- (Wagenschein, 1968, S. 101).

Diese Forderungen stoßen bis heute auf große Zustimmung in der fachdidaktischen Literatur (s. z.B. Sommer, 2005; Labudde, 2006; Buck & Mackensen, 2006; Mikelskis, 2007; Theilmann, 2008a, 2008b).

Phänomene stellen wichtige Ausgangspunkte von Unterricht dar, weil sich aus ihrer Beobachtung und genauen Analyse ein *qualitatives* Verstehen des in Frage stehenden Sachverhaltes entwickeln kann (s. dazu Plöger, 2001). Bei diesem Vorgehen können bereits einige Einflussgrößen als mögliche Variablen in den Beobachtungsfokus rücken und auf einfachem Wege manipuliert werden. Die Lehrperson sollte darauf achten, dass die *qualitativen* Momente dabei deutlich hervortreten, damit daraus dann Hypothesen über die zugrunde liegenden Zusammenhänge aufgestellt werden können, die die SchülerInnen anschließend im Experiment überprüfen (s. oben). Die ermittelten Messdaten sind *quantitativer* Natur, die durch entsprechende grafische Darstellungen leichter interpretiert werden können und gleichzeitig erste Hinweise auf die Mathematisierung der Ergebnisse liefern. Am Ende dieses (Forschungs-)Prozesses steht eine physikalische Formel, die den entsprechenden physikalischen Sachverhalt quantitativ abbildet. Diese (teilweise hoch) abstrakte Formalisierung ist aber nur auf der Basis der vorangegangenen qualitativen Erfassung verstehbar. Lehrpersonen kommt daher die Aufgabe zu, die qualitative und quantitative Erschließung des physikalischen Sachverhalts komplementär zu arrangieren; denn die Effektivität ihres Physikunterrichts hängt „entscheidend davon ab, ob qualitatives und quantitatives Vorgehen gut ausbalanciert sind“ (Labudde, 2006, S. 256).

### Historische Genese physikalischen Wissens

An Lehrpersonen wird die Anforderung gestellt, dass sie im Unterricht „reflektiertes Wissen über ihre Fächer (Metawissen) einsetzen und auf wichtige ideengeschichtliche und wissenschaftstheoretische Konzepte zurückgreifen“ (KMK, 2013, S. 3) können. Im Hinblick auf das Fach Physik sollten sie auch „die Ideengeschichte ausgewählter physikalischer Theorien und Begriffe sowie den Prozess der Gewinnung physikalischer Erkenntnisse (Wissen über Physik)“ (ebd., S. 50) kennen (s. Höttecke, 2007). Diese von der Kultusministerkonferenz formulierte „historische Dimension“ hat eine lange Tradition im physikdidaktischen Denken, die vor allem Martin Wagenschein immer wieder betonte:



Ein notwendiges Element der wissenschaftlichen wie der pädagogischen Ausbildung ist das Vertrautsein mit der geschichtlichen Entwicklung der Physik, vor allem mit ihren Anfängen ... Das gelingt nicht durch einen ‚Überblick‘ der Physik-Geschichte, sondern durch das Studium ausgewählter Quellen ... Physik muss als gewordene und werdende Auseinandersetzung des Menschen mit der Natur vom Lehrer gesehen und gelehrt werden.  
(Wagensein, 1965, S. 382f.)

Beispielsweise lassen sich im Physikunterricht durch den Vergleich der Atomvorstellungen von Joseph J. Thomson (1856–1940) und dessen Schüler Ernest Rutherford (1871–1937) (s. dazu ausführlich Plöger, 2011), eine Vielzahl physikalisch wie wissenschaftstheoretisch bedeutsamer Erkenntnisse bei SchülerInnen anregen. In *physikalischer* Hinsicht ist das Atommodell Thomsons durch zwei zentrale Annahmen bestimmt: Erstens geht Thomson von der *räumlichen* Nähe der gleichmäßig im Raum verteilten positiven und negativen Ladungen aus. Zweitens „unterstellt“ er ihnen eine *statische* Anordnung. Rutherford postulierte das Gegenteil, denn seine Konstruktion des Atoms ist dadurch geprägt, dass nicht die räumliche Nähe und gleichmäßige Verteilung der Ladungen sowie die Statik der Teilchen, sondern die *große Distanz* der unterschiedlichen Ladungen und die *Dynamik* der Elektronen den Aufbau des Atoms bestimmen.

Aus *wissenschaftstheoretischer Sicht* war entscheidend, dass die von Rutherford in seinen Streuversuchen gemachten Beobachtungen im Widerspruch zu den Annahmen der Thomsonschen Theorie standen und mit dieser nicht erklärbar waren: Die auf eine Goldfolie geschossenen positiven Ladungsteilchen hätten nach der Thomsonschen Theorie unabgelenkt durch die Folie hindurchgehen müssen. Tatsächlich waren aber mehr oder weniger große Ablenkungen (Streuungen) festzustellen. Die Deutung dieser Anomalie gelang Rutherford durch die Konstruktion eines neuen Modells, dessen Erklärungskraft in der Annahme lag, dass in allen Atomen nahezu die gesamte Masse der positiven Ladungen in einem Zentrum, dem Kern des Atoms, vereint ist und die weitaus leichteren Elektronen in großer Entfernung um den Kern herumkreisen. Diese neue physikalische Konstruktion stellte einen großen Erkenntnisfortschritt dar, weil sie im Gegensatz zur Vorgängertheorie Thomsons ein größeres Erklärungs- und Prognosepotenzial hatte.

Wenn nun das Rutherfordmodell im Unterricht besprochen wird, müssen im Unterrichtsverlauf auch dessen Grenzen durch weitere Differenzierungen bestimmt werden. Durch dieses Vorgehen können theoretische Standpunkte und Konstrukte kontinuierlich relativiert werden, sodass sich die Wahrheitsfrage angemessener als die Frage nach der Passung des betreffenden Modells auf die empirischen Daten stellt und die distanzierte Betrachtung historisch längerer Zeiträume sowohl die jeweiligen Paradigmenwechsel als auch den Vorläufigkeitscharakter physikalischer Wissensgenerierung deutlich machen kann.

### **Produktiver Umgang mit Fehlern**

In den zuvor erläuterten fachlichen bzw. fachdidaktischen Kategorien ist bereits indirekt die Bedeutung des Vorwissens der SchülerInnen erwähnt worden, das zum Teil im Unterricht erworben, zum Teil aber auch aus dem Alltag mit in den Unterricht gebracht wird (s. z.B. Duit, 1993, 1995a; Jung, 1986). Dieses Vorwissen stellt einen wichtigen Faktor im Lernprozess der SchülerInnen dar (s. z.B. Duit, 1995a, S. 910; Duit & Treagust, 1998; Wiesner 1995), da es das Fundament bildet, aus dem heraus „das neue Wissen konstruiert werden muß“ (Duit, 1995a, S. 910). Es kann aber auch ein „Lernhemmnis“ (ebd.) darstellen. Denn im Vergleich zu dem künftig noch zu erwerbenden fachlichen Wissen weist das

relevante Alltagswissen der SchülerInnen mehr oder weniger große „Defizite“ in Form von Fehlvorstellungen auf, die den Lernerfolg im Verlauf des Unterrichts, ganz besonders aber in Einstiegs- und Erarbeitungssituationen (s. Spychiger, Oser, Hascher, & Mahler, 1999), entsprechend beeinträchtigen können.

Da SchülerInnen letztlich nur auf der Basis ihres Vorwissens neue Sinnzusammenhänge erschließen können (s. z.B. Ausubel, 1968; Ausubel et al., 1980; Jung, 1986; Tergan, 1986), sollten aus *lerntheoretischer* Sicht solche Fehlvorstellungen im Unterricht explizit angesprochen werden. Aber auch aus *motivationspsychologischer* Sicht können diese Fehlvorstellungen nicht einfach übergangen werden, wenn man voraussetzt, dass SchülerInnen aus ihrem Vorwissen heraus für sie selbst *ernsthafte* (wenn auch ggf. unangemessene) hypothetische Zugriffe auf das Phänomen entwickeln und dass „die subjektive Verarbeitung von publik gemachten Fehlern je nach erfolgter *Attribution* durch den Betroffenen sich unterschiedlich“ (Rollett, 1999, S. 78) gestalten kann (s. auch Weinert, 1999, S. 105).<sup>78</sup>

Es ist bisher nur ansatzweise untersucht worden, wie Lehrpersonen mit solchen Fehlern bzw. Fehlvorstellungen im Unterricht produktiv umzugehen wissen. Im Rahmen der IPN-Videostudien konnte aber gezeigt werden, dass Fehler und Fehlvorstellungen im Physikunterricht kaum sichtbar werden bzw. nur bedingt angesprochen werden. Dabei wurde auch nachgewiesen, dass in Extremfällen die Atmosphäre sogar zwischen „merkbar angespannt“ und „positiv“ pendeln kann (s. Seidel, Prenzel, Rimmel, Dalehefte et al., 2006, S. 806f.). Im Zusammenhang damit könnten die zusätzlichen Befunde zur „Lernbegleitung“ auf Fehlervermeidungsstrategien seitens der Lehrperson verweisen, denn bei deren Fragen handelte es sich vorwiegend (80%) um Reproduktionsfragen, selten (5%) wurde mit den Fragen eine Verknüpfung verschiedener physikalischer Sachverhalte (deep reasoning) verlangt. „Sachlich-konstruktive oder positiv-unterstützende Rückmeldung kamen in nur 12 Prozent der Rückmeldungssituationen vor“ (ebd., S. 806).

Über diesen fachspezifischen Befund hinaus konnte Dweck (1996) nachweisen, dass lern- und leistungsorientierte Lehrpersonen im Unterricht nur ungenügend zwischen Lern- und Leistungssituationen trennen, sodass SchülerInnen den gesamten Unterricht eher als Leistungssituationen wahrnehmen und somit das Lernen vor allem gering begabter und ängstlicher SchülerInnen deutlich gehemmt werden kann.

Daher sollten Lehrpersonen nicht nur zwischen Lern- und Leistungssituationen trennen, sondern auch eine „Fehlerkultur“ in ihrem Unterricht etablieren. Im positiven Fall kennt die Lehrperson „den Lerngegenstand sehr gut und ist mit den typischen, wiederkehrenden Fehlern vertraut; sie ist im konkreten Unterricht auf die entsprechenden Fehlersituationen vorbereitet und kann auf eine dem Lernprozeß förderliche Weise damit umgehen“ (Spychiger et al., 1999, S. 59).

---

78 Duit (1993) weist daraufhin, dass diese Alltagsvorstellungen konträr zum physikalisch korrekten Sachverhalt liegen können. Problematisch bei ihrer Richtigstellung ist allerdings, dass die vorliegenden Alltagsvorstellungen nicht einfach als „falsch“ deklariert werden können, da sie „Realität“ der SchülerInnen sind, denn diese Vorstellungen haben „sich vielmehr als Interpretationsrahmen alltäglicher Erfahrungen bestens bewährt“ (Duit, 1993, S. 5). Folglich ist dieses Wissen in einer gewissen Form immunisiert bzw. lösungsresistent (begrenzte Überzeugungskraft empirischer Ergebnisse) und bestimmt die Wahrnehmung entsprechender Phänomene (WahrnehmungsfILTER) (Duit, 1989, 1993).

### **Lernen als konstruktive Tätigkeit von SchülerInnen**

Mit den oben angestellten Überlegungen zum konstruktivistischen Wissenschaftsverständnis von Physik (s. S. 97f.) sind indirekte Auffassungen für das Lehren und Lernen im Unterrichtsfach Physik verbunden. Zunächst ist festzuhalten, dass die Erkenntnis physikalischer Sachverhalte als eigenständige, selbsttätige Konstruktion aufzufassen ist, die jeder Schüler bzw. jede Schülerin individuell erbringen muss:

Verständnisvolles Lernen ist ein aktiver individueller Konstruktionsprozess, in dem Wissensstrukturen verändert, erweitert, vernetzt, hierarchisch geordnet oder neu generiert werden. Auch verständnisvolles rezeptives Lernen ist in diesem Sinne eine Konstruktionsleistung ... Entscheidend für verständnisvolles Lernen ist die aktive mentale Verarbeitung, die sich in der handelnden Auseinandersetzung mit der sozialen oder natürlichen Umwelt ... vollzieht.

(Baumert & Köller, 2000, S. 273f.)

Verschiedene empirische Arbeiten deuten aber daraufhin, dass Lehrpersonen tendenziell weniger dieser konstruktivistischen Auffassung folgen, sondern stärker transmissiv orientiert sind (s. im Überblick z.B. Handal, 2003; aber auch Baird & Mitchell, 1986; Baird & Northfield, 1982). Dabei wird das Lehren als „Weitergabe von Wissen an die Schülerinnen und Schüler [verstanden], die mehr oder weniger passive Rezipienten sind“ (Voss, Kleickmann, Kunter, & Hachfeld, 2011, S. 238, Anm. MK). Im COACTIV-Projekt konnte durch Kombination von Lehrer- und Schülerdaten gezeigt werden, dass sich „transmissive Überzeugungen von Lehrkräften ... als nachteilig für die Unterrichtsqualität und den Lernerfolg der Schülerinnen und Schüler“ erwiesen; „konstruktivistische Überzeugungen standen dagegen in einem positiven Zusammenhang mit Unterrichtsqualität und Lernerfolg“ (ebd., S. 250). Im TEDS-M-Projekt konnten ähnliche Zusammenhänge nachgewiesen werden. Da es sich dort aber um einen internationalen Vergleich handelt, sind die Ergebnisse differenzierter im Hinblick auf zugrunde liegende kulturelle Prägungen und Überzeugungen der Lehrkräfte zu interpretieren, wobei in individualistisch geprägten Ländern eher eine konstruktivistische, in kollektivistisch orientierten Ländern eher eine transmissive Auffassung vom Lehren zu registrieren war (s. Schmotz, Felbrich, & Kaiser, 2010, S. 281, S. 297ff.).

In Entsprechung zu den Befunden aus der Mathematikdidaktik konnten auch für den Bereich des naturwissenschaftlichen Unterrichts entsprechende Evidenzen erbracht werden. Zur Beurteilung konstruktivistisch orientierter Lernumgebungen im *Biologieunterricht* konnten Marsch, Hartwig und Krüger (2009) ein sehr differenziertes Kategoriensystem entwickeln und mittels Videostudien den Zusammenhang zwischen den Überzeugungen von Lehrkräften und ihrem Unterrichtshandeln belegen. In den IPN-Videostudien konnten für den *Physikunterricht* nur bedingt konstruktivistisch orientierte Lernumgebungen nachgewiesen werden (s. z.B. Duit, 2006; Seidel, Prenzel, Rimmele, Schwindt et al., 2006; Seidel, Prenzel, Rimmele, Dalehefte et al., 2006; Widodo & Duit, 2004, 2005). Es wurde beobachtet, dass Lehrpersonen ihren Unterricht eher „stofforientiert“ ausrichten und die Weitergabe des Inhalts als Mittelpunkt ihrer Überzeugungen ansahen. Dieses transmissive Lehrverständnis wirkt sich jedoch eher hinderlich auf die SchülerInnenleistungen aus, da es zu entsprechenden Einschränkungen beim Erwerb physikalischen Wissens führt (Widodo & Duit, 2004, S. 250).

Vor diesem Hintergrund lässt sich mit Blick auf die notwendigen Kompetenzen von Lehrpersonen festhalten: Sie sollten sich weniger als (reine) Wissensvermittler definieren, sondern eher als Moderatoren von Lernprozessen der SchülerInnen verstehen.<sup>79</sup> Mit dieser Forderung wird eine Metaperspektive beschrieben, die dem *gesamten* Lehr-Lernprozess seine grundlegende Ausrichtung vorschreibt. Denn diese basale Kompetenz der Unterrichtsmoderation weist allen anderen Kompetenzen ihren Stellenwert zu, sodass sich die Gestaltung der unterrichtlichen Aktivitäten von Lehrpersonen, die prinzipiell *konstruktivistisch* orientiert sind, deutlich von den Maßnahmen von Lehrpersonen unterscheiden, die eine *transmissive* Orientierung vertreten. Unterschiede treten vor allem bei der Berücksichtigung des Vorwissens der SchülerInnen auf, bei der Auswahl geeigneter Medien und Materialien, bei der Gestaltung des Interaktionsprozesses und dem Austausch von Ideen, bei der Auswahl entsprechender Anwendungsmöglichkeiten der erworbenen Kompetenzen sowie bei der Initiierung der geistigen Selbsttätigkeit und kognitiven Aktivierung (s. dazu im Detail Widodo & Duit, 2004, S. 238ff., 250).

Demzufolge wird mit dieser siebten Kategorie „Lernen als konstruktive Tätigkeit des Schülers“ eine fächerübergreifende lernpsychologische Sicht auf Lehr-Lernprozesse im Physikunterricht ausgelegt, durch die nicht das Lehren, sondern das *Lernen* als entscheidende Größe in den Fokus der *Analyse von Unterricht* rückt: Die SchülerInnen müssen die notwendigen kognitiven Strukturen selbst konstruieren (s. Jung, 1986), aber die Lehrperson hat die Funktion, diesen Konstruktionsprozess durch geeignete didaktisch-methodische Realisierungen zu unterstützen.

### 3.2.4 Kategorien für fachliches (incl. fachdidaktisches) Wissen

Die bisherigen Ausführungen dienten dazu, die Kategorien auf theoretischem Wege zu legitimieren. Sie stellen einen *Minimalkonsens* dar, der durch theoretische wie empirische Arbeiten gedeckt ist (s. KMK, 2004a, 2004b, 2013). Im Folgenden sollen nun die konkreten Kategorien vorgestellt werden, die im Rahmen der Qualitativen Inhaltsanalyse zum Einsatz gebracht werden und mit denen die Interviews der ExpertInnen als auch der NovizInnen unter fachlichen (incl. fachdidaktischen) Gesichtspunkten analysiert werden sollen. Dieses Vorgehen ist analog zur Vorstellung der didaktisch-methodischen Kategorien (s. Kap. 3.2.5), d.h. jede Kategorie wird prägnant definiert, beschrieben und an einem Ankerbeispiel veranschaulicht.

---

79 Insbesondere bei der Planung von Unterricht sollten Lehrpersonen die SchülerInnenvorstellungen (incl. Vorwissen) berücksichtigen, denn nur so können sie den Unterricht an das Niveau der SchülerInnen anpassen und die Inhalte entsprechend didaktisch aufbereiten, damit „sie von den Schülern verstanden werden“ (Duit, 2006, S. 21) können (s. auch Kattmann, Duit, Gropengießer, & Komorek, 1997; Wiesner, 2006). Der Lehrer ist somit „nicht Überbringer des Wissens ... sondern als Anreger zur Eigenkonstruktion“ (Duit, 1993, S. 7) von Wissen für SchülerInnen zu sehen.

## Physikalische Begriffsbildung

**Tab. 24:** Kategorie Physikalische Begriffsbildung

Kategorie	Definition/Beschreibung/Ankerbeispiel
Physikalische Begriffsbildung (Kürzel: PB)	<p><i>Definition:</i> Die Kategorie „Physikalische Begriffsbildung“ bezieht sich auf die Aufgabe des Physikunterrichts, SchülerInnen in die physikalische Fachsprache einzuführen, damit sie konkrete Naturphänomene mit Hilfe abstrakter Begriffe theoretisch deuten und messen können.</p> <p><i>Beschreibung:</i> Physikalische Begriffe repräsentieren konkrete Wirklichkeit in Form mentaler Modelle, bilden Wirklichkeit dabei aber nicht ab, sondern konstruieren sie durch Herausstellen wesentlicher Elemente bzw. kritischer Attribute. Ein kritisches Attribut des Begriffes Lichtbrechung ist z.B. die Grenzfläche zwischen zwei unterschiedlich dichten optischen Medien wie Luft und Wasser, weil an genau diesem Punkt die Brechung stattfindet. Im Prozess der Begriffsbildung müssen die relevanten Begriffe hervorgehoben werden, d.h. die SchülerInnen sollen jene Begriffe lernen, die der Lehrer klar und verständlich einführt. Die in Frage kommenden Begriffselemente müssen einerseits identifiziert und somit isoliert, andererseits aber auch in eine spezifische Beziehung zueinander gesetzt werden, durch die dann der jeweilige Begriff als kognitive Struktur entsteht (Aebli, 1993a). Solche begrifflichen Strukturen stehen immer in Beziehung zu weiteren Begriffen, wodurch mehr oder weniger umfangreiche Begriffsnetze entstehen können. Das Begriffsnetz Lichtbrechung entsteht z.B. durch die Beziehung zwischen den (Teil-)Begriffen Dichte eines optischen Mediums, Einfallslot, Einfallswinkel, Brechungswinkel, zum Lot hin, vom Lot weg. Ein solches Begriffsnetz zu bilden, heißt zugleich, einen bestimmten Sachverhalt zu verstehen. Damit dieser Begriffsbildungsprozess im Unterricht gelingt, sollte die Lehrperson einige wichtige Strategien beachten: Neue Begriffe sollten in der Regel im Kontext realer Phänomene erarbeitet werden (Mikelskis-Seifert &amp; Duit, 2010; Muckenfuß, 2004). Die SchülerInnen müssen dabei die relevanten Elemente nach Möglichkeit selbst verbalisieren, weil die sprachliche Bezeichnung Aufmerksamkeit fokussiert und den Bewusstseitsgrad steigert. Im Sinne Wagenscheins (1968) haben zunächst qualitative Aspekte Vorrang; erst wenn die qualitativen Beziehungen zwischen Elementen einsichtig geworden sind, macht es Sinn, Überlegungen zur Messbarkeit des Phänomens anzustellen. Oft ist es erforderlich, eine ganze Reihe von neuen Begriffen einzuführen (das ist auch bei der zu analysierenden Unterrichtsstunde zur Erarbeitung des Brechungsgesetzes der Fall); SchülerInnen benötigen dann hinreichend Zeit, die Begriffe aufzunehmen. Begriffe, die zunächst keine unmittelbare Bedeutung für den zu erarbeitenden Sachverhalt haben, sondern eher zu dessen sukzessiver Differenzierung dienen, sollten erst dann eingeführt werden, wenn sich die zentralen Begriffe einigermaßen konsolidiert haben. Schließlich ist es wichtig, SchülerInnen immer wieder aufzufordern, die eingeführten Begriffe präzise und konsistent anzuwenden. Das gilt umso mehr für die Lehrperson, die im Hinblick auf präzise Begriffsbildung und -anwendung Vorbildfunktion übernehmen sollte.</p> <p><i>Ankerbeispiel:</i> „Der Lehrer trägt eine Fülle an Begrifflichkeiten an die SchülerInnen heran, teilweise überlappend, parallel oder kurz hintereinander. Dabei vergisst er aber, die neuen Begriffe, die die SchülerInnen noch nicht kennen, zu definieren. Anstatt eines klaren, präzisen Tafelbildes, mit dem die SchülerInnen weiterarbeiten könnten, verliert er sich in diffusen Formulierungen über Reflexion.“</p>

## Durcharbeiten von Fachbegriffen

**Tab. 25:** Kategorie Durcharbeiten von Fachbegriffen

Kategorie	Definition/Beschreibung/Ankerbeispiel
Durcharbeiten von Fachbegriffen (Kürzel: DFB)	<p><i>Definition:</i> Diese Kategorie „Durcharbeiten von Fachbegriffen“ bezieht sich auf die Vertiefung und Konsolidierung neu erarbeiteter Begriffe. Sie müssen durchgearbeitet (Aebli, 1993a) werden, damit die spezifischen Relationen zwischen den Begriffselementen klarer werden bzw. sich ausschärfen.</p> <p><i>Beschreibung:</i> Nach ihrer erstmaligen Erarbeitung sind Begriffe noch relativ instabil, weil die SchülerInnen noch zu sehr am Beispiel „kleben“, die konkreten Merkmale des Kontextes, in dem die Begriffe eingeführt wurden, also noch nicht hinreichend von den wesentlichen, abstrakten Merkmalen unterscheiden können. Ein Transfer des betreffenden Begriffes auf einen anderen Kontext gelingt dann – wenn überhaupt – nur mit Mühe. Zwischen Erarbeitung und Transfer muss deshalb erst ein Durcharbeiten (Aebli, 1993a) des Begriffes erfolgen, damit die Beziehungen zwischen den Begriffselementen klarer hervortreten. Auf diese Weise werden die begrifflichen Strukturen dekontextualisiert, sodass eine Beweglichkeit im Denken und Handeln entsteht, die dann auch den Transfer des Begriffes auf andere Situationen ermöglicht.</p> <p><i>Ankerbeispiel:</i> „Diese letzte Phase erschließt sich mir nicht. Er hat zwar versucht, einen Anwendungskontext zu schaffen, aber es ist nicht klar, worauf er mit dieser Linse hinauswill. In der Unterrichtsstunde geht es um Brechungsgesetze, daher ist der Sprung von der Zeichnung einer Linse zur Fresnel-Linse überhaupt nicht plausibel für die SchülerInnen. Er hat sich aus meiner Sicht zu viel vorgenommen und hätte zum Abschluss der Stunde den SchülerInnen besser die Möglichkeit gegeben, die bisherigen Erkenntnisse zu sichern und abzuspeichern.“</p>

### Hinweise zur Verwendung der Kategorien:

Die Anwendung der Kategorie „Durcharbeiten von Fachbegriffen“ bezieht sich ausschließlich auf die *letzte Phase* des zu analysierenden Unterrichts. Die Demonstration der Fresnel-Linse muss bei den SchülerInnen den Eindruck erwecken, als sei sie ein typisches Anwendungsbeispiel für das Brechungsgesetz. Das ist aber nicht bzw. nur bedingt der Fall: Die Fresnel-Linse ist zunächst einmal eine Linse wie jede andere, weil auch sie einfallendes Licht bricht und dadurch entsprechende Bilder erzeugen kann. In dieser Hinsicht zeigt sie eine Gemeinsamkeit zu der von den SchülerInnen im Experiment verwendeten Linse. In diesem Experiment ging es im Kern um die Bestimmung des Verhältnisses zwischen Einfallswinkel und Brechungswinkel und die Erkenntnis eines gesetzmäßigen Zusammenhangs zwischen diesen Winkeln.

Die Fresnel-Linse behandelt der Lehrer aber im Hinblick auf ihre *technische Gestaltung*: Sie ist so gebaut, dass alles Material, in dem der Strahlengang des Lichtes keine Brechung erfährt, entfernt wird. Nur an den Stellen, an denen das Licht wieder austritt, wird es gebrochen. Thematisiert wird hier also *primär* die Ersparnis des Materials einer Linse bei gleichzeitiger Gewährleistung ihrer Funktion.

Insgesamt gesehen, wird durch die Demonstration und Erklärung der Linse nicht der Begriff des Brechungsgesetzes konsolidiert; das erkennt man auch daran, dass der Lehrer auf

diesen Begriff völlig verzichten kann. Insofern stellt diese letzte Phase einen Bruch zum Vorangegangenen dar.

Es ist wünschenswert, dass die ProbandInnen diese hier erläuterte Inkonsistenz im Vorgehen der Lehrperson erkennen und zur Sprache bringen. Vermutlich wird das nur Personen gelingen, die „vom Fach“ sind. Deshalb ist es auch akzeptabel, Textstellen zu kodieren, in denen zumindest ein *Bruch* im Stundenaufbau erkannt wird, der den SchülerInnen keine klare thematische Orientierung ermöglicht.

*Nicht akzeptabel* ist dagegen eine positive Würdigung der letzten Phase als *gelungene Anwendung* des zuvor Gelernten.



## Experimentieren als zentrale Fachmethode

**Tab. 26:** Kategorie Experimentieren als zentrale Fachmethode

Kategorie	Definition/Beschreibung/Ankerbeispiel
Experimentieren als zentrale Fachmethode (Kürzel: EZF)	<p><i>Definition:</i> Diese Kategorie beschreibt die Methode des Experimentierens als wechselseitigen Prozess von Theorie und Empirie. Dieser Prozess umfasst mehr als die bloße praktische Ausführung des Experiments selbst; denn dieses ist eingebettet in theoretische Vorüberlegungen, Planungen und Dateninterpretationen, stellt also nur einen Teil dieses Gesamtprozesses dar.</p> <p><i>Beschreibung:</i> Experimentieren ist ein mehrstufiger Prozess, der durch spezifische Fragen und Problemstellungen in Gang kommt. Daraus entwickeln sich theoriegeleitete Hypothesen, deren Überprüfung ein begründetes Versuchsdesign erfordert. Die Ausführung des Experiments selbst und die dadurch gewonnenen Messdaten lassen dann Rückschlüsse auf die Angemessenheit der Hypothesen zu. Dieser Prozess spiegelt sich im Unterricht in einer stimmigen Folge von entsprechenden Phasen: Klare Problemstellung, Aufstellen und argumentatives Begründen von Hypothesen, Planung, Durchführung und Auswertung des Experiments. Die über diesen Prozess angestrebte Qualität des physikalischen Wissens hängt in hohem Maße davon ab, dass die einzelnen Phasen auf SchülerInnenseite zu präzisen und kohärenten Lernleistungen führen (Verstehen des Problems, Identifizieren möglicher Variablen, Begründen von Hypothesen, Entwerfen eines geeigneten Experiments, Durchführen des Experiments mit Blick auf die erwarteten Variablenzusammenhänge, Auswerten der ermittelten Daten unter der leitenden Problem- bzw. Fragestellung). In summa zielen diese Teilleistungen darauf, dass auf Seiten der SchülerInnen zunächst ein gedanklicher Lösungsprozess initiiert wird, der darin mündet, ein tiefes Verständnis davon zu entwickeln, was sie mit dem jeweiligen Experiment erkennen wollen.</p> <p><i>Ankerbeispiel:</i> „Jetzt sollte der Versuch protokolliert werden. Die erste Zeile des Versuchsprotokolls wird von ihm vorgegeben. Ich meine, die SchülerInnen in dem Stadium müssten alleine die Versuchsskizze anfertigen können, müssten wissen, dass ein Versuchsprotokoll strukturiert ist nach Versuchsaufbau, -durchführung, -beobachtung und -ergebnis. Sie können das alleine entwickeln. Das fände ich auch die bessere Variante im Sinne des Kompetenzerwerbs, dass man eben nicht nur das Sachwissen erwerben will, sondern tatsächlich auch die Fähigkeit, sich physikalischen Problemen zu nähern und die experimentell zu erschließen.“</p>

### Hinweise zur Verwendung der Kategorien:

Bei der Kodierung ist darauf zu achten, dass mind. 2 Merkmale/Phasen (klare Problemstellung, Aufstellen und argumentatives Begründen von Hypothesen, Planung, Durchführung und Auswertung des Experiments) benannt werden. Das bedeutet, dass die einzelnen aufgezählten Phasen nicht allesamt wörtlich genannt sein müssen, aber es muss deutlich werden, dass die analysierende Person das Experiment nicht als bloße praktische Betätigung identifiziert und positiv bewertet, sondern die Notwendigkeit sieht, es in einen stimmigen Denk- und Handlungsprozess zu integrieren.

**(weitere) fachliche Arbeitsweisen**

**Tab. 27:** Kategorie (weitere) fachliche Arbeitsweisen

Kategorie	Definition/Beschreibung/Ankerbeispiele
(weitere) fachliche Arbeitsweisen (Kürzel: FAW)	<p><i>Definition:</i> Diese Kategorie beschreibt (über das Experimentieren als zentrale Fachmethode hinaus) weitere fachliche Arbeitsweisen als typische Operationen, die die Erarbeitung des jeweiligen Sachverhaltes in spezifischer Weise ermöglichen. Im Fach Physik gehören dazu vor allem: Phänomene beobachten und genau analysieren; Messdaten grafisch darstellen; kritisches Beurteilen von Messwerten; qualitative Zusammenhänge (ansatzweise) mathematisieren.</p> <p><i>Beschreibung:</i> Die hier genannten physikalischen Arbeitsweisen treten insbesondere im Zusammenhang mit den Teilprozessen des Experimentierens auf (also z.B. bei der Planung, Durchführung und Auswertung). Die Arbeitsweisen dienen dabei als Mittel zum Zweck; das Beurteilen von gemessenen Werten lässt z.B. Rückschlüsse auf die angemessene Variablenkontrolle im Experiment zu. Um diese instrumentelle Funktion zu erfassen, werden weitere fachliche Arbeitsweisen hier als eine eigenständige Kategorie geführt und von der Kategorie „Experimentieren als zentrale Fachmethode“ deutlich unterschieden. Während bei dieser Kategorie nach Hinweisen auf die Stufen des Experimentierprozesses zu suchen ist (Problemstellung, Hypothesenbildung, Versuchsplanung, ...), ist bei jener darauf zu achten, ob die Lehrperson eine der oben genannten Arbeitsweisen als effektives Mittel zum Gelingen des Experimentierprozesses einsetzt.</p> <p>Grundlegend für die fachlichen Arbeitsweisen ist die Orientierung am Phänomen, d.h. ausgehend vom sinnlich Wahrnehmbaren sollen jene spezifisch physikalischen Arbeitsweisen den Schülern ermöglichen, den Sachverhalt zu erarbeiten.</p> <p><i>Ankerbeispiele:</i></p> <p>a) Phänomene beobachten und genau analysieren: „Das kann man natürlich machen, an Alltagserfahrungen anknüpfen. Jeder hat das vielleicht schon einmal gesehen. Aber wie geht er dann mit den Alltagserfahrungen der SchülerInnen um? Er bewertet sie direkt, sie passen ihm anscheinend nicht in den Kram. Das heißt: Sie werden zurückgedrängt. Und er kommt dann im Prinzip mit ganz anderen Beispielen. Also da würde ich sagen, die Hinführung zum Phänomen ist so nicht besonders gut gelungen.“</p> <p>b) Messdaten grafisch darstellen: „Die Messdaten lässt er in eine Tabelle eintragen, aber wieso werden diese nicht grafisch dargestellt? Zum einen würden die SchülerInnen beim Betrachten der Grafik sehen, dass das am Anfang einigermaßen linear ist und dann abflacht. Und zum anderen lernen sie, eine Grafik zu interpretieren.“</p> <p>c) kritisches Beurteilen von Messwerten: „Die Auswertung der Messdaten ist eine Katastrophe und naturwissenschaftlich auch völlig unsauber. Wenn ich Werte erhoben habe, dann schaue ich mir die erstmal an und falls die abweichen sollten von dem, was in der Wissenschaft gängig ist, dann muss ich darüber eine Fehlerdiskussion führen.“</p> <p>d) qualitative Zusammenhänge mathematisieren: „Die Frage des Lehrers darf eigentlich gar nicht heißen ‚sind die Messwerte proportional‘, sondern ‚Ist ein proportionales Modell sinnvoll oder passend‘. Das sind in den Naturwissenschaften ganz zentrale Dinge.“</p>

**Hinweise zur Verwendung der Kategorien:**

Die Kategorie „(weitere) fachliche Arbeitsweisen“ muss deutlich von der Kategorie „Experimentieren als zentrale Fachmethode“ abgegrenzt werden. Ausgehend von sinnlich wahrnehmbaren Phänomen sollen die SchülerInnen darin geschult werden, jene fachlich spezifischen Arbeitsweisen der Physik kennenzulernen bzw. anzuwenden. Der Experimentierprozess soll bei den SchülerInnen einen gedanklichen Problemlöseprozess initiieren, der durch die jeweiligen fachlichen Arbeitsweisen unterstützt bzw. überprüft wird.

## Historische Genese physikalischen Wissens

**Tab. 28:** Kategorie Historische Genese physikalischen Wissens

Kategorie	Definition/Beschreibung/Ankerbeispiel
Historische Genese physikalischen Wissens (Kürzel: HGPW)	<p><i>Definition:</i> Diese Kategorie bezieht sich auf den Sachverhalt, dass physikalisches Wissen geschichtlich gewachsen ist und insofern den jeweiligen Entwicklungsstand physikalischer Theorien zu bestimmten Zeiten widerspiegelt.</p> <p><i>Beschreibung:</i> Ein hinreichendes Wissen um die historische Genese ist für Lehrpersonen in mehrfacher Hinsicht von Bedeutung. Die unterschiedlichen Deutungen von Phänomenen zu bestimmten Zeiten (z.B. Wärme, Elektrizität usw.) führen zu der wissenschaftstheoretisch bedeutsamen Erkenntnis, dass Begriffe und Theorien immer perspektivische Konstruktionen des betreffenden Sachverhaltes darstellen. In der distanzierten Betrachtung längerer Zeiträume werden der Wandel solcher Perspektiven und damit der Vorläufigkeitscharakter der Wissensgenerierung deutlich. Für den Physikunterricht ist die historische Perspektive u.a. deshalb wichtig, weil sich die Entwicklung physikalischer Vorstellungen bei SchülerInnen in gewisser Weise mit der geschichtlichen Entwicklung physikalischen Denkens parallelisieren lassen. So ist z.B. die geschichtlich überholte Vorstellung von Wärme als stoffliche Substanz auch in den Alltagsvorstellungen von SchülerInnen wiederzufinden. Das Nachstellen historischer Situationen kann daher eine durchaus produktive Auseinandersetzung anregen, weil SchülerInnen ihr Alltagswissen relativ leicht an solche Situationen anbinden können.</p> <p><i>Ankerbeispiel:</i> „Der Lehrer versucht, die SchülerInnen zurückzusetzen in eine Zeit vor vierhundert Jahren. Diesen historischen Ansatz zu wählen ist durchaus legitim und in den Naturwissenschaften ein geeigneter Weg, weil die SchülerInnen mit ihrem Alltagswissen genau da stehen, wo letzten Endes auch die Wissenschaftler vor Entdeckung des Gesetzes standen.“</p>

### Hinweise zur Verwendung der Kategorien:

Im zu analysierenden Unterricht kündigt die Lehrperson zwar einen historischen Rückblick bzw. das experimentelle Nachstellen einer historischen Forschungssituation an, geht aber kurz darauf einen völlig anderen Weg der Erarbeitung. Es kommt bei dieser Kategorie darauf an, dass die ProbandInnen diesen angekündigten historischen Weg als prinzipiell realisierbar und geeignet beurteilen. Dazu müssen nicht unbedingt entsprechende Begründungen abgegeben werden; hinreichend wäre also bereits eine Aussage über die Angemessenheit eines solchen historischen Ansatzes. Als *unangemessen* sind allerdings jene Aussagen zu bewerten, die das historisch orientierte Arbeiten aus unterschiedlichsten Gründen ablehnen (z.B. „SchülerInnen haben kein Interesse an Geschichte.“).

## Produktiver Umgang mit Fehlern

**Tab. 29:** Kategorie Produktiver Umgang mit Fehlern

Kategorie	Definition/Beschreibung/Ankerbeispiel
Produktiver Umgang mit Fehlern (Kürzel: PUF)	<p><i>Definition:</i> Die Kategorie „Produktiver Umgang mit Fehlern“ erfasst die Reaktionen der Lehrperson auf SchülerInnenbeiträge, die der zu erarbeitenden Sache (zunächst) nicht angemessen sind.</p> <p><i>Beschreibung:</i> Sachfehler treten beim sinnstiftenden Lernen mehr oder weniger zwangsläufig auf. Beim Erwerb physikalischen Wissens erklärt sich diese Tatsache insbesondere aus dem hypothetisch-deduktiven Vorgehen; die für die Lösung eines physikalischen Problems generierten Hypothesen stehen solange in Frage, bis das Experiment darüber entschieden hat.</p> <p>Die Aufstellung von Hypothesen ist allerdings eine Anforderung, die vom zur Verfügung stehenden Vorwissen abhängt. Ist das Wissen brüchig, nicht wirklich verstanden oder nicht hinreichend konsolidiert, lassen sich keine sachangemessenen Hypothesen entwickeln. Im Fach Physik kommt hinzu, dass die im Alltag erworbenen begrifflichen Konzepte denen der Physik nicht entsprechen und deswegen ein Konzeptwechsel notwendig wird, der aber nur durch große Lernanstrengungen gelingen kann.</p> <p>Solche Wissensdefizite manifestieren sich im SchülerInnenverhalten als „Fehler“. Sie lassen sich in der Regel schnell korrigieren, wenn bereits Gelerntes zum Teil in Vergessenheit geraten ist. Hier reichen oft kurze Wiederholungsphasen aus, um Vergessenes in Erinnerung zu rufen. Im kreativen Prozess der Hypothesenbildung spielen allerdings auch die außerunterrichtlich erworbenen Alltagskonzepte eine entscheidende Rolle, weil die SchülerInnen darauf oft als Argumentationsbasis zurückgreifen (müssen). Die Inkongruenz dieser Konzepte mit den physikalischen führt dann zwangsläufig zu „Fehlern“.</p> <p>Da die Alltagskonzepte für die SchülerInnen hohe Plausibilität haben, ist es völlig unangebracht, sie als Fehler abzustempeln. Entsprechende Schüleraussagen müssen daher ernst genommen und als Anlass für weitere produktive Denkprozesse genutzt werden (Widersprüche aufdecken, kognitive Konflikte provozieren, empirische Fakten mit vorgetragenen Argumenten konfrontieren usw.). Zusammenfassend bedeutet das also für die Kategorie „Produktiver Umgang mit Fehlern“, dass die ProbandInnen in der Analyse des Unterrichts danach fragen müssen, inwiefern die Lehrperson „Fehler“ der SchülerInnen für das weitere Lernen nutzt.</p> <p><i>Ankerbeispiel:</i> „Das Alltagswissen der SchülerInnen wird vom Lehrer belächelt; ob die Aussage mit dem Fischen im Wasser oder das Glas aus Fasern besteht. Das sind Alltagserfahrungen, die er eigentlich überhaupt nicht gebrauchen kann, aber eine Behauptung wie die, das Glas aus Fasern beseht, ist typisch für SchülerInnenvorstellungen, die diese mit in den Unterricht bringen. Solche Vorstellungen muss man aufgreifen und in den Unterricht integrieren, ansonsten wird das Lernen blockiert.“</p>

**Hinweise zur Verwendung der Kategorien:**

Fehler bzw. Fehlvorstellungen seitens der SchülerInnen gehören zum alltäglichen Unterricht und müssen konstruktiv genutzt werden, um sachadäquates Wissen aufzubauen. Guter Unterricht zeichnet sich daher durch eine gewisse Fehlertoleranz und den Aufbau einer entsprechenden Fehlerkultur aus. Die Kategorie „Produktiver Umgang mit Fehlern“ berücksichtigt zwei entscheidende Aspekte: Einerseits müssen SchülerInnen Fehler machen bzw. Fehlvorstellungen äußern (dürfen) und andererseits müssen Fehler von der Lehrperson identifiziert, thematisiert und korrigiert werden. Für die Kodierung von Textstellen folgt daraus, dass auch negative Formulierungen der interviewten Personen kodiert werden müssen, wenn diese damit bspw. das Übergehen oder Ignorieren solcher Fehlvorstellungen thematisieren.

## Lernen als konstruktive Tätigkeit von SchülerInnen

**Tab. 30:** Kategorien Lernen als konstruktive Tätigkeit von SchülerInnen

Kategorie	Definition/Beschreibung/Ankerbeispiel
Lernen als konstruktive Tätigkeit von SchülerInnen (Kürzel: LKS)	<p><i>Definition:</i> Die Kategorie „Lernen als konstruktive Tätigkeit von SchülerInnen“ beschreibt die Aneignung von Wissen als einen aktiven Prozess der Schülerin bzw. des Schülers, durch den nur er bzw. sie selbst den physikalischen Sachverhalt konstruieren kann bzw. muss.</p> <p><i>Beschreibung:</i> „Lernen als konstruktive Tätigkeit von SchülerInnen“ ist eine allgemeine Perspektive auf den Lernprozess der SchülerInnen. Diese Kategorie zielt auf eine globale lernpsychologische Sicht, durch die nicht das Lehren, sondern das Lernen als entscheidende Größe in den Fokus der Analyse von Unterricht rückt: Es ist der Schüler bzw. die Schülerin, der bzw. die die notwendigen kognitiven Strukturen bilden/konstruieren muss. Die Funktion der Lehrperson besteht demnach darin, diesen Konstruktionsprozess durch geeignete didaktisch-methodische Realisierungen zu unterstützen. Diese konstruktivistische Sicht von Lernen wendet sich gegen eine naive „Vermittlungsperspektive“, die der irrigen Annahme folgt, wo gelehrt wird, wird auch gelernt.</p> <p><i>Ankerbeispiel:</i> „Was ich bisher gesehen habe, ist vor allen Dingen Instruktion, aber für Konstruktion von Seiten der SchülerInnen ist da gar kein Platz und was unterrichtlich in den Köpfen der SchülerInnen stattfindet, bekommt er mit seiner dozierenden Unterrichtsweise nicht mit. Mögliche Irrwege, die durchaus lerneffektiv sein können, versucht er durch den sicheren Weg des Vortrags zu umgehen. Er weiß, wo die Reise hingeht und die SchülerInnen vollziehen das nach. Es ist insofern ein nachdenkender Unterricht, bei dem der Lehrer vor denkt und die SchülerInnen das Gedachte nachdenken.“</p>

### Hinweise zur Verwendung der Kategorien:

Diese Kategorie muss von der didaktisch-methodischen Kategorie „Aktivierung und Motivierung zur geistigen Selbsttätigkeit (kognitive Aktivierung)“ *scharf* unterschieden werden. Die Kategorie „Aktivierung und Motivierung zur geistigen Selbsttätigkeit (kognitive Aktivierung)“ bezieht sich auf *konkrete* Situationen und Handlungen, in denen die Selbsttätigkeit (nicht oder nur ungenügend) in Gang kommt, oder auf Alternativvorschläge, wie diese Selbsttätigkeit *angeregt* werden könnte. Die Kategorie „Lernen als konstruktive Tätigkeit von SchülerInnen“ zielt dagegen nicht auf einzelne Situationen und Handlungen, sondern charakterisiert eine *grundsätzliche Haltung* der Lehrperson gegenüber der inneren Lerntätigkeit der SchülerInnen, die über die gesamte Unterrichtsstunde hinweg erkennbar wird.



### 3.3 Kategorien für die formale Qualität des Wissens (Verstehenstiefe)

In den vorangegangenen Teilkapiteln 3.1 und 3.2 wurden die Kategorien zur Erfassung des didaktisch-methodischen und des fachlichen (incl. fachdidaktischen) Wissens hergeleitet. In diesem Teilkapitel möchte ich die Kategorien für die *formale* Qualität des Wissens erläutern und begründen. Diese formale Qualität des Wissens werde ich über den Begriff der „Verstehenstiefe“ operationalisieren.

Den Ausgangspunkt der Beschreibung der formalen Qualität des Wissens stellt das fünfstufige Modell der Analysekompetenz von Plöger und Scholl (2014a) dar, das in Kapitel 1 vorgestellt wurde (s. Kap. 1.4.2). Dieses Modell wurde von ihnen für die quantitative Messung der Analysekompetenz angehender und praktizierender Lehrpersonen entwickelt und der Prüfung der faktoriellen Struktur des Konstrukts Analysekompetenz mittels Konfirmatorischer Faktorenanalysen zugrunde gelegt. Um es für den Zweck meiner Qualitativen Inhaltsanalyse nutzen zu können, bedarf es allerdings einer gewissen *Transformation*. Diese Transformation werde ich zunächst beschreiben, um daran anschließend die einzelnen Kategorien für „Verstehenstiefe“ mit ihren Definitionen, Beschreibungen und Ankerbeispielen vorzustellen.

#### 3.3.1 Das Kompetenzmodell von Plöger und Scholl als Bezugsrahmen

Das fünfstufige Modell der Analysekompetenz von Plöger und Scholl wurde bereits ausführlich in Kapitel 1 erläutert (s. Abb. 4, S. 36). Deshalb seien an dieser Stelle lediglich die wichtigsten Annahmen, die dem Modell zugrunde liegen, noch einmal betont:

- Die Unterrichtsanalyse stellt einen Prozess der Informationsverarbeitung dar, bei dem Informationen unterschiedlicher Komplexität zu verarbeiten sind.
- Der Grad der jeweils notwendigen Komplexität der Informationsverarbeitung wird über verschiedene Stufen hinweg abgebildet (*ansteigende Komplexität*).
- Dabei ist jede Stufe des Modells durch entsprechende Analyseleistungen gekennzeichnet, die erfüllt sein müssen, damit eine Person dieser Stufe zugeordnet werden kann. Leistungen einer jeweils höheren Stufe können nur dann erbracht werden, wenn die betreffende Person auch fähig ist, Anforderungen niedrigerer Stufen gerecht zu werden. Systematisch gesehen bedeutet das, dass komplexere Anforderungen (= höhere Stufen des Modells) weniger komplexere Anforderungen (= niedrigere Stufen des Modells) organisieren.
- Das Modell besteht aus drei Hauptstufen (Analytische Kompetenz, Synthetische Kompetenz und Prozesskompetenz) und fünf Unterstufen.
- Auf den Stufen 1 und 2 (= *Analytische* Kompetenz) sind ProbandInnen in der Lage, *einzelne* Situationen und/oder Handlungen zu registrieren (Stufe 1) und deren graduelle Ausprägung (Stufe 2) zu bestimmen. Sowohl solche einzelnen Situationen bzw. Handlungen als auch deren Ausprägung sind sinnlich wahrnehmbar; sie stehen somit für das, was in der Diskussion als „Sichtstrukturen von Unterricht“ (s. z.B. Aebli, 1971; Kliehne, 2006; Oser & Baeriswyl, 2002; Oser & Patry, 1994; Reyer, 2004; Seidel, 2003a, 2003b) bezeichnet wird. Um Leistungen, die mit Stufe 1 und 2 beschrieben werden, erbringen zu können, ist durchaus auch Informationsverarbeitung notwendig, da es sich

um eine begrifflich geleitete Wahrnehmung bzw. Registrierung von Unterrichtsereignissen handelt. Von „echten“ Verstehensprozessen, über die ich im Folgenden die formale Qualität von Wissen zu operationalisieren versuche, kann auf diesen beiden Stufen aber noch keine Rede sein. Denn Verstehen setzt die Herstellung bzw. das Erkennen von *Zusammenhängen* voraus. Die Wahrnehmung einzelner Situationen bzw. Handlungen (incl. ihrer Ausprägungen) wird aber diesem Kriterium nicht gerecht.

- Das Verstehen von entsprechenden Zusammenhängen zwischen (mindestens zwei) Situationen und/oder Handlungen wird im fünfstufigen Modell durch die Stufen 3 bis 5 beschrieben. Diese Stufen bilden Analyseleistungen ab, die nicht allein durch direkte Wahrnehmung zustande gekommen sind, sondern durch einen (zusätzlichen) Akt der Konstruktion bzw. Interpretation. Dadurch werden unterrichtliche Zusammenhänge erschlossen, die in der Literatur gewöhnlich als „Tiefenstrukturen von Unterricht“ (s. z.B. Aebli, 1971; Klieme, 2006; Oser & Baeriswyl, 2002; Oser & Patry, 1994; Seidel, 2003b; Seidel & Prenzel, 2006, 2007; Reyer, 2004) bezeichnet werden.
- Stufe 3 ist dadurch definiert, dass ProbandInnen einen Zusammenhang zwischen Situationen bzw. Handlungen und deren Folgen/Effekten herstellen können. Auf Stufe 4 sind die ProbandInnen in der Lage, größere didaktisch-methodische *Sinneinheiten* zu bilden und mögliche Alternativen zu diesen anzugeben. Beide Stufen beschreiben somit Leistungen, durch die die Personen einen Zusammenhang zwischen verschiedenen Situationen herstellen müssen. Deshalb benennen Plöger und Scholl diese Stufen auch als *Synthetische Kompetenz*.
- Stufe 5 beschreibt dagegen die *Prozesskompetenz*. Personen, deren Analysekompetenz diesen Level erreicht, haben den *gesamten* Unterrichtsprozess im Blick. Sie sind im Vergleich zu den Stufen 3 und 4 nicht nur in der Lage, mehrere Situationen/Handlungen als eine didaktisch-methodische Sinneinheit zu verstehen, sondern solche Sinneinheiten in den Kontext des gesamten Unterrichtsverlaufs einzuordnen und unter dem Aspekt der Lernwirksamkeit zu beurteilen.
- Mit Blick auf diese fünf Stufen des Modells ist also zusammenfassend festzuhalten, dass es auf dem Grundgedanken der *ansteigenden Komplexität der Informationsverarbeitung* beruht: Zunächst werden nur *einzelne* Situationen/Handlungen erfasst (Stufe 1 und 2), dann mindestens zwei Situationen/Handlungen zueinander in Beziehung gesetzt (Stufe 3 und 4) und schließlich der gesamte Unterrichtsprozess betrachtet (Stufe 5). Die Leistungen der unteren Stufen stellen Voraussetzungen für die darüber liegenden Stufen dar, sodass höhere Stufen die Leistungsanforderungen der niedrigeren beinhalten.

Im Rahmen des quantitativen Validierungsversuchs von Plöger et al. (2015) waren diesen fünf Stufen des Modells entsprechende Testitems zugeordnet (s. Tab. 2, S. 45), deren Lösung in inhaltlicher Hinsicht hinreichendes didaktisch-methodisches bzw. fachliches Wissen erforderte. In formaler Hinsicht war für die Lösung eines Items aber auch ein bestimmter Grad der Komplexität der Informationsverarbeitung erforderlich. Diesem jeweiligen Komplexitätsgrad entsprechend war dann das Item der jeweiligen Stufe des Modells zugeordnet.

Bei der Auswertung der Tests hatten KodiererInnen zu entscheiden, ob die von den ProbandInnen gegebenen Antworten den in einem Kodiermanual zusammengestellten mög-

lichen Antworten entsprachen und somit akzeptiert werden konnten. Eine als richtig akzeptierte Beantwortung eines Items bedeutete dann, dass die Person eine Analyseleistung erbracht hatte, die hinsichtlich der formalen Qualität ihres Wissens (= Komplexität der Informationsverarbeitung) einer bestimmten Stufe des Modells entsprach.

Dieses Prozedere der Auswertung unterscheidet sich in mehrfacher Hinsicht von dem, das in meiner Qualitativen Inhaltsanalyse zur Anwendung kommt. Die Qualitative Inhaltsanalyse ist auf die Auswertung von Interviews ausgerichtet, in denen sich die ProbandInnen relativ frei äußern können, sodass mit hoher Wahrscheinlichkeit eine Vielfalt an Aussagen zustande kommt. In der Auswertung der Interviews müssen die KodiererInnen dann zum einen entscheiden, ob sie eine Textpassage in inhaltlicher Hinsicht als Aktivierung didaktisch-methodischen oder fachlichen (incl. fachdidaktischen) Wissens deuten. Für diesen Zweck verwenden sie die in den beiden vorangehenden Teilkapiteln entwickelten Kategorien. Zum anderen müssen die KodiererInnen auch im Hinblick auf die formale Qualität des Wissens prüfen, für welchen Grad der Komplexität der Informationsverarbeitung die Aussagen der ProbandInnen stehen. Zu diesem Zweck sollen sie die im Folgenden herzuleitenden Kategorien für „Verstehenstiefe“ nutzen.

Diese Kategorien für die „Verstehenstiefe“ stellen eine Transformation der durch die fünf Stufen modellierten Komplexitätsgrade der Informationsverarbeitung dar. Die folgende Tabelle 31 gibt zunächst einen Überblick darüber, wie die Übersetzung der fünf Stufen in verschiedene Level von Verstehenstiefe vorgenommen werden soll.

**Tab. 31:** Stufen des Kompetenzmodells von Plöger et al. (2015) in Entsprechung zu den Aspekten formaler Wissensqualität

Stufen des Kompetenzmodells	Aspekte formaler Wissensqualität, die als Kategorien zur Analyse der Interviews genutzt werden
Stufe 5	Verstehenstiefe – Level 3b
Stufe 5	Verstehenstiefe – Level 3a
Stufe 4	Verstehenstiefe – Level 2b
Stufe 4	Verstehenstiefe – Level 2a
Stufe 3	Verstehenstiefe – Level 1
Stufe 1 und 2	

Den zentralen Bezugspunkt für diese Transformation bilden die in der linken Spalte angeführten fünf Stufen des Modells. In der rechten Spalte sind dann entsprechende Kategorien bzw. Level von Verstehenstiefe als Äquivalente für die verschiedenen Grade der Komplexität der Informationsverarbeitung (und damit für die formale Qualität des Wissens) angeführt. Für die beiden unteren Stufen des Modells werden allerdings keine Äquivalente benannt. Denn wie oben bereits dargelegt, handelt es sich bei diesen beiden untersten Stufen nicht um „echte“ Verstehensleistungen, sondern lediglich um die Registrierung einzelner Situationen bzw. Handlungen, die noch nicht mit weiteren Situationen und/oder Handlungen in Zusammenhang gebracht werden. Verstehen heißt dagegen, um mit Aebli (1993a, 1993b) zu sprechen, spezifische Beziehungen zwischen „Elementen“ herzustellen. Bei der Analyse von Unterricht bilden die jeweiligen Situationen bzw. Handlungen

die betreffenden „Elemente“, zwischen denen Zusammenhänge konstruiert bzw. interpretativ erschlossen werden müssen. Solche Verknüpfungen von Elementen werden aber erst ab Stufe 3 hergestellt, indem (mindestens zwei) Situationen bzw. Handlungen miteinander in Verbindung gebracht werden und auf die Tiefenstruktur des Unterrichts fokussiert wird.

### **3.3.2 Bestimmung der Kategorien für die formale Qualität des Wissens (Verstehenstiefe) in Entsprechung zu den fünf Stufen des Kompetenzmodells von Plöger und Scholl**

#### **Verstehenstiefe – Level 1 (in Entsprechung zu Stufe 3)**

Situationen und Handlungen *und deren Effekte bzw. Folgen* angemessen einschätzen zu können, entspricht der *Stufe 3* des Modells. Um solche Leistung erbringen zu können, ist es notwendig, dass vorausgehende Handlungen mit späteren in Verbindungen gebracht und letztere als *Effekte bzw. Folgen* der vorausgegangen gedeutet werden.

Damit ProbandInnen die „richtige“ Einschätzung der Effekte vornehmen, müssen sie auch die „richtigen“ (spezifischen) Handlungen und Folgen zueinander ins Verhältnis setzen. Das wäre etwa der Fall, wenn die ProbandInnen zu Recht monierten, dass die Lehrperson in der videografierten Stunde keine prägnanten Beispiele als Einstieg wählt und dies zur Folge hat, dass die SchülerInnen kein klares Problemverständnis entwickeln können. Dies könnten die ProbandInnen z.B. auf diese Weise zum Ausdruck bringen: „Anhand der Beispiele wird das Phänomen nicht deutlich, denn die Beispiele überlagern sich teilweise oder lenken sogar von wichtigen Aspekten ab.“ Eine solche Aussage würde dann mit der Kategorie „Verstehenstiefe – Level 1“ kodiert.

#### **Verstehenstiefe – Level 2a (in Entsprechung zu Stufe 4)**

Mit Stufe 4 des Modells werden Leistungen beschrieben, bei denen die ProbandInnen einzelne Handlungen oder Situationen zu größeren didaktisch-methodischen Einheiten verdichten können. Wenn Personen solche didaktisch-methodischen Einheiten erkennen können, sind sie meistens auch in der Lage, deren Lernwirksamkeit einschätzen zu können. Im Falle zu geringer Lernwirksamkeit können sie dann auch *begründbare Alternativen* zu diesen Einheiten angeben und die erforderlichen Lehr- und Lernhandlungen benennen, mit denen sich die vorgeschlagenen Alternativen umsetzen ließen.

Als Indikator für die Kategorie „Verstehenstiefe – Level 2a“ gilt demnach die Angabe *begründeter Alternativen* zu didaktisch-methodischen Einheiten. So könnten z.B. die ProbandInnen feststellen, dass die Lehrperson zu Beginn oder im Verlauf der Stunde das bestehende Vorwissen der SchülerInnen unzureichend aktiviert. Als Alternative könnten sie dann etwa vorschlagen: „Die Lehrperson hätte den SchülerInnen ein entsprechend präpariertes Arbeitsblatt aushändigen können, welches sie mit ihren SitznachbarInnen in fünf Minuten ausfüllen, um das notwendige Vorwissen zum Thema Brechung zu aktivieren.“

### Verstehenstiefe – Level 2b (in Entsprechung zu Stufe 4)

Verschiedene Situationen und/oder Handlungen liegen im Unterrichtsverlauf zeitlich gesehen (mehr oder weniger weit) auseinander. Ob die ProbandInnen zwischen solchen zeitlich getrennten Situationen Verbindungen herstellen können, lässt sich formal gesehen an *Rückverweisen* auf vorangegangene Unterrichtsszenen einerseits und an *Antizipationen* des vermuteten weiteren Unterrichtsverlaufs andererseits festmachen.<sup>80</sup>

Ein typischer *Rückverweis* stellt z.B. die Aussage dar: „Mir fällt auch hier wieder auf, dass die Lehrperson sehr stark doziert und die SchülerInnen zu wenig zum tiefen Nachdenken aktiviert. Das war auch bereits in der vorangegangenen Phase der Fall.“ Eine Äußerung für das Vorliegen einer *Antizipation* wäre z.B.: „Die Lehrperson ist hart an der Grenze zum Spielerischen, man müsste jetzt einmal sehen, wie sich das im weiteren Verlauf darstellt.“ Solche Aussagen (*Rückverweis*, *Antizipation*) indizieren, dass die betreffende Person Zusammenhänge zwischen Situationen erkannt hat, und werden deshalb mit „Verstehenstiefe – Level 2b“ kategorisiert.

### Verstehenstiefe – Level 3a (in Entsprechung zu Stufe 5)

„Verstehenstiefe – Level 3a“ korrespondiert mit Stufe 5 des Modells. Auf diesem Level, ebenso wie auf dem „Verstehenstiefe – Level 3b“ (s. dazu unten), werden Analyseleistungen erbracht, die zeigen, dass die Personen den *gesamten* Unterrichtsprozess unter dem Gesichtspunkt der Lernwirksamkeit einschätzen können. Diese Einschätzung kann eigentlich erst nach Beobachtung des kompletten Unterrichts erfolgen, weil dann alle notwendigen Informationen vorliegen. Aus der Expertiseforschung ist aber bekannt, dass sich ExpertInnen, im Gegensatz zu NovizInnen, schon nach kurzer Zeit ein sehr präzises Bild von einer Situation machen können. Diese Fähigkeit der schnellen und sicheren Einschätzung, die Bromme (1992) als *subitizing* bezeichnet, nutze ich in meiner Qualitativen Inhaltsanalyse zur Operationalisierung der Kategorie „Verstehenstiefe – Level 3a“.

Bezogen auf die Analyse von Unterricht bedeutet *subitizing*, dass ExpertInnen bei der Beobachtung einer Unterrichtsstunde schon nach kurzer Zeit eine präzise (wenn auch bis auf Weiteres hypothetische) Vorstellung von der Qualität der gesamten Stunde entwickeln, obwohl sie den weiteren Verlauf des Unterrichts noch gar nicht kennen. Aus diesem Grunde erfolgt die Präsentation der videografierten Unterrichtsstunde nicht am Stück, sondern aufgeteilt in fünf Abschnitte.

Im Zusammenhang mit dieser Art der sukzessiven Präsentation von Informationen wird dann die Fähigkeit operationalisiert, sich schnell ein hypothetisches Bild vom gesamten Unterrichtsprozess und dessen Lernwirksamkeit machen zu können. Die ProbandInnen müssen bereits zu Beginn der Analyse eine *hypothetische Einschätzung der Realisierbarkeit der Zielstellung*, die dem beobachteten Unterricht unterstellt wird, abgeben. Entsprechende Statements erfordern, dass die ProbandInnen drei Größen miteinander in Verbindung bringen können: Sie müssen (1) eine Vorstellung vom *Ziel der Unterrichtsstunde* entwickeln, (2) das *Vorwissen der SchülerInnen* einschätzen und dann (3) über den Abgleich von Zielsetzung und Vorwissen eine hypothetische Einschätzung der *Realisierbarkeit des Unterrichtsziels* vornehmen.

80 Damit solche Antizipationen oder Rückverweise auch tatsächlich zustande kommen können, wurde das Video in mehrere Abschnitte untergliedert, sodass eine hinreichende Anzahl an Möglichkeiten bestand, um auf den bereits gesehenen Unterrichtsausschnitt rückverweisen zu können oder kommende Aspekte des Unterrichts vorwegzunehmen.

Liegen entsprechende Aussagen der ProbandInnen hinsichtlich dieser drei Größen (Ziel der Unterrichtsstunde; Vorwissen der SchülerInnen; Realisierbarkeit des Unterrichtsziels) vor, so können sie mit der Kategorie „Verstehenstiefe – Level 3a“ kodiert werden. Zeitlich gesehen müssen diese Aussagen direkt nach der Präsentation der ersten, spätestens aber im Anschluss an die zweite Sequenz erfolgen.

### **Verstehenstiefe – Level 3b (in Entsprechung zu Stufe 5)**

Zur Operationalisierung der Kategorie „Verstehenstiefe – Level 3b“ greife ich auf Brommes Beschreibung der Kompetenzen und Funktionen von Lehrpersonen zurück (s. Bromme, 1997). Basierend auf der Auswertung empirischer Arbeiten postuliert Bromme als zentrale Anforderungen an Lehrpersonen die „Aufrechterhaltung einer Struktur von Lehrer- und Schüleraktivitäten“ (S. 189) sowie die damit einhergehende „Entwicklung des Stoffes im Unterricht“ (S. 190) unter effektiver und strukturierter Nutzung der Zeit.

Diese fachbezogenen Anforderungen sind jedoch nicht nur für die Planung und Durchführung von Unterricht von entscheidender Bedeutung, sondern auch für dessen Analyse, in der es darauf ankommt, „eine bestimmte Ordnung zu erkennen und zu konstruieren“ (Bromme, 1992, S. 151). Die im Zusammenhang damit erzeugten „Bedeutungen erlauben die Unterscheidung von Wichtigem und Unwichtigem, jeweils relativ zu den Zielsetzungen und Bedingungen der Anforderung“ (ebd.). Das gelingt nur, wenn die ProbandInnen nach entsprechenden Informationen suchen, um das geschehene Unterrichtsgeschehen unter dem Gesichtspunkt der Lernwirksamkeit evaluieren zu können. Dieses komplexe Zusammenwirken von Informationssuche und -evaluation über den gesamten Unterrichtsprozess hinweg bildet das wesentliche Merkmal der Kategorie „Verstehenstiefe – Level 3b“. Diese Kategorie zielt also darauf festzustellen, ob ProbandInnen ihre antizipierte Beurteilung konsequent über die Analyse des *gesamten* Unterrichts hinweg verfolgen und damit verifizieren oder ggf. falsifizieren.

Im folgenden Teilkapitel stelle ich die Kategorien, wie sie für die Kodierung des Textmaterials genutzt wurden, mit ihren Definitionen, Beschreibungen und Ankerbeispielen vor. Als Äquivalente für diese Stufen 3 bis 5 werden in der Tabelle insgesamt fünf Kategorien verwendet, wobei sowohl auf Stufe 4 als auch auf Stufe 5 jeweils eine zweifache Differenzierung vorgenommen wird. Die Differenzierung innerhalb dieser beiden Stufen (gekennzeichnet jeweils durch a und b) stellt aber keine Hierarchisierung innerhalb dieser Stufen dar, sondern lediglich zwei verschiedene Möglichkeiten, die Verstehenstiefe auf diesem Level zu operationalisieren. Durch diese Differenzierungen ist es dann möglich, ein größeres Spektrum an Aussagen zu kodieren.

Die nähere Bestimmung der fünf Kategorien bzw. Level für Verstehenstiefe soll im folgenden Teilkapitel erfolgen, wobei zu klären ist, was die jeweils definierenden Merkmale sind, um im Prozess der Kodierung klar entscheiden zu können, welche Aussage unter welche Kategorie subsumiert werden kann.

### 3.3.3 Kategorien und Ankerbeispiele für die formale Qualität des Wissens

Die im Folgenden zu bestimmenden Kategorien sollen den Komplexitätsgrad zu verarbeitender Informationen bei der Analyse von Unterricht erfassen. Sie werden in Entsprechung zu den jeweiligen Stufen des Modells der Analysekompetenz (s. Abb. 4, S. 36) in vier Level unterteilt.

#### Verstehenstiefe – Level 1

**Tab. 32:** Kategorie Verstehenstiefe – Level 1

Kategorie	Definition/Beschreibung/Ankerbeispiel
Verstehenstiefe – Level 1 (Kürzel: VT-1)	<p><i>Definition:</i> „Verstehenstiefe – Level 1“ steht für die Fähigkeit, Verbindungen zwischen Situationen oder Handlungen und den dadurch hervorgerufenen Effekten herzustellen. Diese Kategorie bezieht sich auf die Stufe 3 des Kompetenzmodells.</p> <p><i>Beschreibung:</i> Von der Lehrperson arrangierte Situationen oder initiierte Handlungen sind im (problemorientierten) Unterricht letztlich auf Lerneffekte auf Seiten der SchülerInnen ausgerichtet. Solche vermuteten Effekte können unmittelbar, aber auch mit einem gewissen zeitlichen Abstand auf entsprechende Situationen oder Handlungen eintreten. Aussagen auf dieser Stufe beziehen sich auf einzelne Handlungen seitens des Lehrers und der Frage, was bei den SchülerInnen passiert (Effekte).</p> <p><i>Ankerbeispiel:</i> „Er gibt auch unzureichende Erklärungen für die Begründung, warum er diesen halbkreisförmigen Körper verwendet. Das ist für die SchülerInnen nicht nachvollziehbar, weil sie den Körper nicht gesehen haben. Also hier, ich würde mal sagen, der größte Fehler oder die größte Schwäche dieser Phase ist, dass er über Sachen redet, die man leicht sichtbar machen könnte, aber er macht sie nicht sichtbar.“</p>



## Verstehenstiefe – Level 2a

**Tab. 33:** Kategorie Verstehenstiefe – Level 2a

Kategorie	Definition/Beschreibung/Ankerbeispiel
Verstehenstiefe – Level 2a (Kürzel: VT-2a)	<p><i>Definition:</i> „Verstehenstiefe – Level 2a“ bezieht sich auf die Fähigkeit, begründbare Alternativen zu didaktisch-methodischen Sinneinheiten angeben zu können. Diese Kategorie bezieht sich auf die Stufe 4 des Kompetenz-modells.</p> <p><i>Beschreibung:</i> Didaktisch-methodische Sinneinheiten kommen durch Verdichtungen einzelner Situationen bzw. Handlungen zu größeren (Teil-)Einheiten zustande. Wenn eine Person in der Lage ist, solche didaktisch-methodischen Einheiten zu erfassen und angemessen auf Lerneffektivität zu beurteilen, dann kann sie in der Regel auch Alternativen angeben, von denen sie sich eine höhere Lerneffektivität verspricht. Da solche Alternativen als Folgen des (partiellen oder gesamten) Unterrichtsverständnisses anzusehen sind, können sie begründet werden. „Verstehenstiefe – Level 2a“ wird daher gemessen am Vorliegen begründeter Alternativen. Solche Begründungen liegen z.B. nicht vor, wenn ProbandInnen lediglich mehr SchülerInnenaktivierung fordern, aber nicht angegeben wird, mit welchen „Mitteln“ eine höhere SchülerInnenaktivierung zu erreichen wäre.</p> <p><i>Ankerbeispiel:</i> „Die Lehrperson könnte anstatt der Bilder Realphänomene zeigen, z.B. einen Stab ins Wasser halten; der Gegenstand scheint dann geknickt, ist er aber nicht. Das schafft bei den SchülerInnen einen kognitiven Konflikt.“</p>

### Hinweise zur Verwendung der Kategorien:

Bei der Kodierung ist darauf zu achten, dass begründete Alternativen vorliegen, d.h. es muss seitens der Interviewten eine klare Aussage gemacht werden, was anders und warum es anders gemacht werden sollte. Nicht kodiert werden einfache Äußerungen bzw. Feststellungen, wie bspw. „Man müsste den SchülerInnen eine Minute mehr Zeit zum Nachdenken geben.“ Die Aussage stellt zwar einen Alternativvorschlag dar, jedoch fehlt eine *angemessene* Begründung.

**Verstehenstiefe – Level 2b**

**Tab. 34:** Kategorie Verstehenstiefe – Level 2b

Kategorie	Definition/Beschreibung/Ankerbeispiele
Verstehenstiefe – Level 2b (Kürzel: VT-2b)	<p><i>Definition:</i> „Verstehenstiefe – Level 2b“ steht für die Fähigkeit, Verbindungen zwischen Situationen oder Handlungen herzustellen und die dadurch konstruierten Zusammenhänge durch (hypothetische) Vorgriffe (Antizipationen) oder durch Rückverweise explizieren zu können. Diese Kategorie bezieht sich auf die Stufe 4 des Kompetenzmodells.</p> <p><i>Beschreibung:</i> Verbindungen von Situationen oder Handlungen zu größeren Teileinheiten können auf sehr unterschiedliche Weise hergestellt werden. Um festzustellen, ob eine Person den dritten Level erreicht, werden hier zwei Möglichkeiten herausgegriffen, nämlich die Herstellung von Verbindungen durch Antizipationen oder Rückverweise.</p> <p><i>Ankerbeispiele:</i> Antizipation: „Die Lehrperson ist hart an der Grenze zum Spielerischen, man muss nun schauen, wie sich das im weiteren Verlauf darstellt.“ Rückverweis: „Was sich in der Eingangsphase bereits andeutete, setzt sich in dieser leider fort; die Lehrperson ist nicht ernsthaft an SchülerInnenbeiträgen interessiert.“</p>

## Verstehenstiefe – Level 3a

**Tab. 35:** Kategorie Verstehenstiefe – Level 3a

Kategorie	Definition/Beschreibung/Ankerbeispiel
Verstehenstiefe – Level 3a (Kürzel: VT-3a)	<p><i>Definition:</i> „Verstehenstiefe – Level 3a“ bezieht sich auf die Fähigkeit, den gesamten Unterrichtsprozess zu erfassen. Dieser Level korrespondiert mit der Stufe 5 des Kompetenzmodells.</p> <p><i>Beschreibung:</i> Als entscheidendes Indiz hierfür gilt, ob sich die ProbandInnen durch die Analyse der ersten Unterrichtsszene einen Überblick verschaffen können, durch den drei Größen in einen Zusammenhang gesetzt werden: (1) Es muss das Ziel des Unterrichts umrissen werden. (2) Die sich in der Zielstellung ausdrückenden Intentionen müssen zum (vermuteten) Vorwissen der SchülerInnen ins Verhältnis gesetzt werden. (3) Aus dieser Verbindung von Zielen und Vorwissen wird eine (vorerst hypothetische) Einschätzung der Realisierbarkeit der Zielstellung entwickelt. Von diesen drei Größen müssen die ersten beiden bereits im Kommentar der Eingangsszene genannt und in Verbindung gesetzt werden. Die dritte Größe – die Einschätzung der Lerneffektivität – muss ebenfalls genannt sein.</p> <p><i>Ankerbeispiel:</i> „Es gefällt mir gut, wenn die Lehrperson sagt, wir wollen uns nicht nur der Sache selbst zuwenden, sondern auch die Frage stellen, wie wir die Sache angehen können. Ob die SchülerInnen allerdings dafür gerüstet sind, das mache ich ein großes Fragezeichen.“</p>

### Hinweise zur Verwendung der Kategorien:

Die Kategorie „Verstehenstiefe – Level 3a“ ist eine „Globalkategorie“ bei der die interviewte Person eine *holistische* Sicht auf den Unterrichtsprozess nachweist, d.h. sie hat den gesamten Unterrichtsverlauf im Blick. Aufgrund der Komplexität dieser Kategorie müssen bei der Kodierung zwei Aspekte berücksichtigt werden:

- 1) Niedrigere Stufen der Verstehenstiefe können innerhalb von „Verstehenstiefe – Level 3a“ auftreten.
- 2) Das Vorkommen der Kategorie „Verstehenstiefe – Level 3a“ kann (vermutlich) nicht an einer einzigen Textstelle festgemacht werden. Vielmehr ist zu erwarten, dass ein „verstreutes“ Auftreten über das gesamte Interview vorliegt, jedoch immer mit Blick auf die Ziele des Unterrichts, das Vorwissen der SchülerInnen, sowie die Verbindung von Zielen und Vorwissen.

**Verstehenstiefe – Level 3b****Tab. 36:** Kategorie Verstehenstiefe – Level 3b

Kategorie	Definition/Beschreibung/Ankerbeispiel
Verstehenstiefe – Level 3b (Kürzel: VT-3b)	<p><i>Definition:</i> Die Kategorie „Verstehenstiefe – Level 3b“ steht für die fortlaufende Überprüfung des zu Beginn der Analyse hypothetisch unterstellten Unterrichtsverlaufs. Diese Kategorie bezieht sich auf die fünfte Stufe des Kompetenzmodells.</p> <p><i>Beschreibung:</i> Diese Kategorie muss im Zusammenhang mit der Kategorie „Verstehenstiefe – Level 3a“ gesehen werden, durch die die Verbindung zwischen Ziel, Vorwissen und (vermuteter) Erreichbarkeit des Ziels gekennzeichnet wird. Verstehenstiefe – Level 3a ist erreicht, wenn die ProbandInnen diese Verbindung beschreiben kann. Dabei handelt es sich zunächst um eine <i>hypothetische</i> Konstruktion des erwarteten Unterrichtsverlaufs. Von ExpertInnen wäre zu erwarten, dass sie diesen hypothetischen Vorgriff im Verlauf der weiteren Analyse konsequent evaluieren, also verifizieren oder falsifizieren, und dann ggf. modifizieren. Handlungstheoretisch gesehen ist die Kategorie „Verstehenstiefe – Level 3b“ also eine „Kontrollkomponente“ (Dulisch, 1985, S. 77), durch die die von ProbandInnen beobachteten Situationen oder Handlungen samt ihrer Resultate mit der übergreifenden Zielstellung bzw. den anzustrebenden Teilzielen laufend abgeglichen werden. Die Situationen und Handlungen stehen für die Aktivitätsstruktur, die Resultate für die erreichten oder nicht erreichten Lerneffekte.</p> <p><i>Ankerbeispiel:</i> „Das Experiment wird nicht als Lernprodukt ausgewertet und nicht als weitere Lerngelegenheit genutzt; das entspricht seinem gesamten Unterrichtsskript, das auf Sicherheit und Nachvollzug ausgerichtet ist und nicht die Lernmöglichkeiten schafft, wie ich sie mir wünsche. Aber das habe ich eben auch schon einmal angesprochen: Er nimmt die SchülerInnen überhaupt nicht mit, redet über ihre Köpfe hinweg und weiß überhaupt nicht, was die SchülerInnen verstanden haben.“</p>

## 4. Hypothesen

Im vorangegangenen Kapitel wurden die Kategorien vorgestellt, mit denen das Textmaterial (Interviews) kodiert worden ist. Die Ergebnisse dieses Kodierprozesses werden im nächsten Kapitel präsentiert und im Detail erläutert. Um diese Ergebnisse – die zunächst ja nur eine mehr oder weniger große Anzahl an Kodierereignissen für die jeweiligen Kategorien darstellen – unter der übergreifenden Fragestellung meiner Arbeit interpretieren zu können, umreiß ich in diesem Kapitel vorab noch einmal diese Fragestellung und leite aus ihr zwei Generalhypothesen sowie weitere, daraus resultierende Teilhypothesen ab, die die Analyse der Daten leiten.

Diese Hypothesen zielen allesamt auf die Beantwortung der Frage, ob sich die von Plöger et al. (2015) theoretisch postulierte und auf quantitativem Wege bestätigte Struktur des Konstrukts Analysekompetenz auch auf qualitativem Wege replizieren lässt. Über diesen, auf die Struktur des Konstrukts Analysekompetenz zielenden Validierungsversuch hinaus konnten Plöger und Scholl in ihrer Studie (2014a) zeigen, dass die erhobenen Daten sich im Sinne des sogenannten well-known-group-Vergleichs (s. unten) interpretieren lassen. Diese beiden Validierungsstrategien (Validierung des Konstrukts, Validierung per well-known-group-Vergleich) versuche ich auch in meiner qualitativen Untersuchung zur Beurteilung der Validität zu nutzen. Ich umreiß hier noch einmal kurz diese beiden Strategien und deren Realisierung in den Studien von Plöger und Scholl (2014a) sowie Plöger et al. (2015) und formuliere dann die damit verbundenen zwei Generalhypothesen.

- 1) In der Untersuchung von Plöger et al. (2015) sollte der Frage der Validität des Konstrukts Analysekompetenz auf quantitativem Wege nachgegangen werden. Die zentrale Annahme dabei lautete, dass das Konstrukt Analysekompetenz durch zwei Dimensionen, eine inhaltliche und eine formale, bestimmt ist (s. Kap. 1.6.1, S. 41f.). Die inhaltliche Dimension setzt sich aus zwei Faktoren zusammen: dem didaktisch-methodischen und dem fachlichen (incl. fachdidaktischen) Wissen. Die formale Dimension bildet den Grad der für die Analyseleistung notwendigen Komplexität der Informationsverarbeitung ab. Im Rückgriff auf das fünfstufige Kompetenzmodell ist diese formale Dimension durch drei Faktoren definiert: Analytische Kompetenz (Stufen 1 und 2 des Modells), Synthetische Kompetenz (Stufe 3 und 4) und Prozesskompetenz (Stufe 5). Um die durch diese Annahmen postulierte Struktur des Konstrukts zu prüfen, wurden die Dimensionen und zugehörigen Faktoren auf unterschiedliche Weise kombiniert, sodass sich dadurch fünf mögliche Modelle konstruieren ließen. Konfirmatorische Faktorenanalysen (CFA) zeigten (s. Tab. 3, S. 46), dass sich hinreichende Fit-Werte nur dann zeigten, wenn das betreffende Modell die postulierte zweidimensionale Struktur des Konstrukts (incl. aller Faktoren) spiegelte.
- 2) Der zweite Validierungsbeleg von Plöger und Scholl (2014a) basierte auf dem sogenannten well-known-group-Vergleich (Cronbach & Meehl, 1955, DeVellis, 1991, Hattie & Cooksey, 1984, Rubin & Babbie, 2015). Dessen Grundgedanke zielt auf den Vergleich zweier Gruppen, von denen man vermutet, dass das in Frage stehende Merkmal (hier Analysekompetenz) bei den Personen der einen Gruppe besonders hoch und bei den Personen der anderen Gruppe besonders niedrig ausgeprägt ist. Entsprechen die

Testergebnisse diesen erwarteten Unterschieden, so ist dies ein wichtiges Indiz für die Validität des zugrunde liegenden Konstrukts Analysekompetenz. Die von Plöger und Scholl (2014a) erhobenen Daten zeigten, dass Fach- und SeminarleiterInnen (ExpertInnen) durchweg die höchsten Stufen der Analysekompetenz erreichten, die Studierenden (NovizInnen) dagegen nur die niedrigeren Stufen. Zwischen diesen beiden Gruppen lagen die Analyseleistungen der ReferendarInnen und praktizierenden Lehrpersonen (s. Kap. 1.5.2, S. 40 sowie Plöger und Scholl, 2014a, S. 104). Diese großen Unterschiede zwischen den ExpertInnen und den NovizInnen waren mit Blick auf die Erkenntnisse der Expertiseforschung „erwartbar“ und ließen sich somit als weiteren Beleg für die Validität des zugrunde liegenden Konstrukts Analysekompetenz interpretieren.

In Analogie zu diesen beiden Validierungsstrategien (Konstruktprüfung, well-known-group-Vergleich) war auch die Konzeption meiner Studie von vornherein methodisch darauf angelegt, die faktorielle Struktur des Konstrukts Analysekompetenz zu überprüfen (s. Kap. 1) und zudem „signifikante“ Unterschiede in den Analyseleistungen von ExpertInnen und NovizInnen sichtbar zu machen (s. Kap. 2).

An diese beiden Validierungsstrategien sind zwei generelle Hypothesen geknüpft:

- **Generalhypothese 1:** Mit der Qualitativen Inhaltsanalyse kann die von Plöger, Scholl und Seifert nachgewiesene zweidimensionale Struktur (incl. Faktoren) des Konstrukts Analysekompetenz repliziert werden.
- **Generalhypothese 2:** Die Qualitative Inhaltsanalyse repliziert die signifikanten Unterschiede zwischen ExpertInnen und NovizInnen.

Diese beiden generellen Hypothesen werden im Folgenden durch weitere Teilhypothesen differenziert. Denn nach Bortz und Döring (2006) wird die Konstruktvalidität desto höher, „je mehr gut gesicherte Hypothesen ihre Überprüfung bestehen“ (Bortz & Döring, 2006, S. 201). Im Rahmen dieser Teilhypothesen wird über die theoretisch postulierte Struktur des Konstrukts hinaus noch ein weiterer Aspekt berücksichtigt, nämlich der *fachliche* Hintergrund der ProbandInnen. Plöger und Scholl (2014a) hatten zeigen können, dass Personen, die das Fach Physik bzw. ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichteten bzw. studierten, signifikant bessere Analyseleistungen erbrachten als Personen, die keines dieser Fächer vertraten.

Um auch dieses Ergebnis überprüfen zu können, waren daher in meinem Sampling Personen vertreten, die sich hinsichtlich ihrer Fachzugehörigkeit unterschieden (s. Kap. 2.3.4, S. 57f.). Mit Blick auf diese Differenzierungen ergeben sich folgende Teilhypothesen:

- a) Hypothesen hinsichtlich der *inhaltlichen* Dimension von Analysekompetenz
  - H1: ExpertInnen aktivieren bei der Analyse des Unterrichts didaktisch-methodische Kategorien in umfangreicherem Maße als NovizInnen.
  - H2: ExpertInnen greifen bei der Analyse des Unterrichts in größerem Umfang auf fachliche (incl. fachdidaktische) Kategorien zurück als NovizInnen.
- b) Hypothesen hinsichtlich der *formalen* Dimension von Analysekompetenz
  - H3: ExpertInnen weisen mehr Nennungen hinsichtlich der Verstehenstiefe von Unterricht auf als NovizInnen.
  - H4: NovizInnen kommen nicht über „Verstehenstiefe – Level 2b“ hinaus.

- H5: Die Komplexität der Wahrnehmung erstreckt sich bei ExpertInnen über alle Level der Verstehenstiefe.
- c) Hypothesen hinsichtlich des fachlichen Hintergrundes der ProbandInnen
- H6: ExpertInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten, verwenden häufiger fachliche (incl. fachdidaktische) Kategorien als ExpertInnen, die diese Fächer nicht unterrichten.
  - H7: NovizInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik studieren, aktivieren häufiger fachliche (incl. fachdidaktische) Kategorien als NovizInnen, die diese Fächer nicht studieren.
  - H8: ExpertInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten, erbringen eine tiefere Analyse des Unterrichts (höhere Level der Verstehenstiefe) als ExpertInnen, die diese Fächer nicht unterrichten.
  - H9: NovizInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik studieren, erzielen eine tiefere Analyse des Unterrichts (höhere Level der Verstehenstiefe) als NovizInnen, die diese Fächer nicht unterrichten.

Diese neun Hypothesen werden durch den Vergleich der Kategorienvorfälle der ExpertInnen und NovizInnen geprüft. Um über diese Hypothesen hinaus zwei weitere (siehe zu H10 und H11, S. 134) prüfen zu können, werde ich die qualitativ ermittelten Daten quantitativ weiterverarbeiten, um auf diese Weise Signifikanzprüfungen, Reliabilitätsanalysen sowie Trennschärfe- und Korrelationsberechnungen durchführen zu können. Quantifizierungen im Sinne von Häufigkeitsanalysen sind im Rahmen einer Qualitativen Inhaltsanalyse durchaus gebräuchlich (s. Grunenberg & Kuckartz, 2010; Kuckartz, 2016; Mayring, 2015), die von mir hier genutzten Analysen werden allerdings eher selten durchgeführt (s. z.B. Fakis, Hilliam, Stonely, & Townend, 2014).

Zu diesem Zweck orientiere ich mich an den Begrifflichkeiten für Reliabilitätsanalysen, wie sie z.B. bei der Bewertung von Testitems durchgeführt werden (s. z.B. Fisseni, 1997). Testitems sollten insbesondere eine hinreichende Reliabilität und eine möglichst hohe Korrelation mit anderen Items der betreffenden Skala aufweisen. Während die Reliabilität die Messgenauigkeit der Items spiegelt, deuten hohe Korrelationen darauf hin, dass die Items eine homogene Skala bzw. Teilskala bilden.

Um die erhaltenen Kodierungen quantitativ weiterverarbeiten zu können, wird jede einzelne kodierte Aussage als ein „Treffer“ gewertet, für den – ähnlich wie bei einer richtigen Antwort eines Testitems – ein Punkt vergeben wird. Jedes Kodierereignis steht somit für die richtige (d.h. von den KodiererInnen akzeptierte) Anwendung von Wissen und stellt einen quantitativen Ausdruck der Wissensaktivierungen dar. Differenzen, die durch unterschiedliche Zahlen bzw. Summen repräsentiert werden, können somit als entsprechende Differenzen hinsichtlich der Ausprägung der Wissensaktivierung interpretiert werden.

Durch diese quantitative Betrachtung der Kodierereignisse werden die genutzten Kategorien als Variablen betrachtet. Demnach beinhaltet das in Kapitel 3 eingeführte Kategoriensystem in der inhaltlichen Dimension zwölf Variablen für den Faktor des didaktisch-methodischen Wissens und sieben Variablen für den Faktor des fachlichen (incl. fachdidaktischen) Wissens. In der formalen Dimension umfasst der Faktor Verstehenstiefe (verschiedene Grade der Komplexität, mit der Informationen verarbeitet werden können) fünf



Variablen. Diese drei Faktoren lassen sich – in Analogie zu Itemskalen – als Skalen bzw. Teilskalen eines Konstrukts (hier Analysekompetenz) interpretieren, die ein Set definierter Variablen enthalten.

Die spezifische Ausprägung dieser Variablen ergibt sich daraus, wie oft eine Person Wissen aktiviert, das einer bestimmten Variablen zugeordnet werden kann. Wenn die Variablen einer Skala inhaltlich homogen sind, ist davon auszugehen, dass ein und dieselbe Person ähnlich hohe Ausprägungen auf allen Variablen dieser Teilskala zeigt. So ist z.B. zu erwarten, dass eine Person mit hohem didaktisch-methodischen Wissen nicht nur auf einigen wenigen Variablen dieser Teilskala hoch scoret, sondern auch auf allen anderen.

Die so ermittelten Wissensaktivierungen stellen intervallskalierte Daten, auf deren Basis die Reliabilität der Skala berechnet werden soll, um so beurteilen zu können, ob die durch die Qualitative Inhaltsanalyse ermittelten Kodierungen als ausreichend genaue „Messungen“ über alle Variablen interpretiert werden können.

Darüber hinaus sollen die vermuteten Unterschiede zwischen ExpertInnen und NovizInnen mithilfe des Mann-Whitney-Tests (U-Test) auf statistische Signifikanz geprüft werden. Ob diese Ergebnisse der Signifikanzprüfung auch eine praktische Bedeutsamkeit besitzen, wird über eine entsprechende Effektstärkenberechnung zu klären sein (s. z.B. Pospeschill, 2006).

Unter Berücksichtigung dieser Ausführungen ergeben sich folgende Teilhypothesen:

d) Hypothesen hinsichtlich der statistischen Berechnungen

- H10: Die vorzunehmenden Reliabilitätsanalysen belegen eine hinreichend genaue Messung des didaktisch-methodischen und fachlichen (incl. fachdidaktischen) Wissens in inhaltlicher Hinsicht sowie eine hinreichend genaue Erfassung der Verstehenstiefe in formaler Hinsicht.
- H11: Es zeigen sich signifikante Unterschiede zwischen den ExpertInnen und NovizInnen zum einen hinsichtlich der Aktivierung didaktisch-methodischen und fachlichen (incl. fachdidaktischen) Wissens und zum anderen bezüglich der in der Analyse sichtbaren Verstehenstiefe.

## 5 Ergebnisdarstellung

Im vorangegangenen Kapitel wurden die Hypothesen meiner Untersuchung vorgestellt. In diesem Teilkapitel möchte ich nun mit Blick auf diese Hypothesen die Ergebnisse der Qualitativen Inhaltsanalyse darstellen, kommentieren und damit die übergreifende Forschungsfrage meiner Arbeit nach der faktoriellen Struktur des Konstrukts Analysekompetenz beantworten.

### 5.1 Überprüfung der Inter- und Intracoderreliabilität

Bevor die Ergebnisse des Kodierverfahrens vorgestellt werden, soll zunächst über das erzielte Maß an Inter- und Intracoderreliabilität, das im Prozess der Kodierung erreicht wurde, berichtet werden. Diese beiden Maße stehen für die *Reproduzierbarkeit* der Kodierungen und für die *Intersubjektivität* des methodischen Vorgehens (s. Kap. 2.5.4, S. 66ff.). Das praktische Verfahren zur Ermittlung der Inter- und Intracoderreliabilität sei hier noch einmal kurz in Erinnerung gerufen: Als Basis dienten die von vier Personen im Konsensverfahren erzielten Kodierungen. Für die Prüfung der Reproduzierbarkeit und Intersubjektivität dieser Kodierungen sollten anschließend zwei externe Kodiererinnen<sup>81</sup>, also Personen, die nicht an diesem Verfahren beteiligt waren, die entsprechenden Textstellen auf die gleiche Weise kodieren, wie das im ersten Kodierdurchgang der Fall war. Mit Inter- und Intracoderreliabilität ist dabei gemeint (s. z.B. Kuckartz, 2016; Lissmann, 2008; Mayring, 2015):

- *Intracoderreliabilität* bedeutet, dass eine Kodiererin ein und dieselbe Textstelle zu unterschiedlichen Zeitpunkten mit derselben Kodierung versieht.
- Mit der *Intercoderreliabilität* werden die Kodierungsentscheidungen (mind.) zweier Kodiererinnen miteinander verglichen. Dabei sollten beide Kodiererinnen für ein und dieselbe Textstelle dieselbe Kategorie vergeben.

Zur Ermittlung der Intercoderreliabilität überprüfen die beiden Kodiererinnen ca. 10% aller markierten Textstellen, insgesamt also 72 Textstellen. Zur Bestimmung der Intracoderreliabilität wurde eine Kodiererin gebeten, die Textstellen nach fünf Wochen erneut zu kodieren.

Darüber hinaus wurde in einem abschließenden Schritt geprüft, ob die Kodierentscheidungen der beiden externen Kodiererinnen mit den im konsensuellen Verfahren festgestellten Kodierungen übereinstimmen. Diese „Validitätsprüfung“ (Früh, 2015, S. 189) entspricht dem *traditionellen* Verständnis von Validität und wird über die Frage definiert, „inwieweit der Test das misst, was er messen soll“ (Rost, 2004, S. 33). Wenn die Kodierungen der beiden externen Kodiererinnen mit den im Konsensverfahren ermittelten Kodierungen

---

<sup>81</sup> Da es sich um zwei weibliche Personen handelte, die die Kodierungen vornahmen, wird hier, in Entsprechung zu Kapitel 2 (s. Kap. 2.5.4, S. 67), ebenfalls die weibliche Form gewählt.

übereinstimmen, ist dies über die Reliabilität hinaus ein Indiz für die Validität der Kodierungen, denn dann kann angenommen werden, dass die externen Kodiererinnen genau das getan haben, was die am Konsensverfahren beteiligten Personen realisiert haben: Sie teilen ein *gemeinsames* Verständnis des Kategoriensystems, d.h. der Definitionen und Beschreibungen der Kategorien samt zugehöriger Ankerbeispiele (s. *Prüfung der Validität*, Kap. 2.5.5, S. 69f.).

Die Ergebnisse der Reliabilitäts- und Validitätsprüfungen sind in Tabelle 37 zusammengefasst:

**Tab. 37:** Überprüfung der Text-Kategorien-Zuordnung

Faktoren des Konstrukts Analysekompetenz		Intracoder- reliabilität	Intercoder- reliabilität	Prüfung der Validität
Inhaltliche Dimension	Didaktisch-methodisches Wissen	.73	.77	.74
	Fachliches (incl. fachdidak- tisches) Wissen	.74	.75	.73
Formale Dimension	Verstehenstiefe	.71	.74	.76

In der linken Spalte sind die Faktoren der inhaltlichen (didaktisch-methodisches und fachliches Wissen) und der formalen Dimension (Verstehenstiefe) angeführt. Als Zielwert wurde zur Beurteilung der Intra- und Intercoderreliabilität, in Anlehnung an Fleiss & Cohen (1973), ein Kappa-Koeffizient von .6 bis .75 angestrebt. Da sämtliche Werte im und teilweise über dem anvisierten Zielbereich (.6 bis .75) liegen, ist davon auszugehen, dass die Wiederholbarkeit, die Stabilität und somit die *Zuverlässigkeit* der Text-Kategorien-Zuordnung gegeben ist. Vor dem Hintergrund dieser guten Werte ist zudem eine wichtige Voraussetzung geschaffen, die Kodierereignisse mittels Reliabilitätsanalysen zu überprüfen (s. dazu weiter unten).

## 5.2 Ergebnisse der Qualitativen Inhaltsanalyse

In diesem Teilkapitel werden die Häufigkeiten der Wissensaktivierungen von ExpertInnen und NovizInnen dargestellt und miteinander verglichen (s. Kap. 5.2.1). Daran anschließend wird der fachliche Hintergrund berücksichtigt, um zu sehen, ob sich Unterschiede zwischen Personen zeigen, die das Fach Physik, ein anderes naturwissenschaftliches Fach und/oder Mathematik oder keines dieser Fächer unterrichten bzw. studieren (s. Kap. 5.2.2).

### 5.2.1 Unterschiede zwischen ExpertInnen und NovizInnen

In der folgenden Tabelle 38 sind die (vorläufigen)<sup>82</sup> Hauptergebnisse der Qualitativen Inhaltsanalyse dargestellt. Im linken Teil der Tabelle sind die Faktoren der inhaltlichen Dimension (didaktisch-methodisches Wissen und fachliches Wissen) und der formalen Dimension (Verstehenstiefe) aufgeführt. Die Zellen der folgenden Spalten stellen die *Summen* der kategorisierten Textstellen (Kodierereignisse) dar, die in der jeweiligen Gruppe (ExpertInnen vs. NovizInnen) für den betreffenden Faktor der inhaltlichen oder formalen Dimension zustande gekommen sind.

Jedes einzelne Kodierereignis ist ein Indiz dafür, dass die ProbandInnen eine spezifische Wissensfacette aktiviert haben und diese auch verbalisieren konnten. So sind beispielsweise in den Interviews der ExpertInnen 223 Textstellen als Aktivierung didaktisch-methodischen Wissens und 124 Textstellen als Aktivierung fachlichen Wissens kodiert worden. Zudem konnten für die ExpertInnen 134 Stellen identifiziert werden, die für den Grad der jeweiligen Verstehenstiefe stehen. In den Interviews der NovizInnen fanden sich 136 Wissensaktivierungen für didaktisch-methodisches Wissen und 55 Aktivierungen des fachlichen (incl. fachdidaktischen) Wissens. Des Weiteren wurden 36 Textstellen ausfindig gemacht, die den Grad der Verstehenstiefe abbilden.

In der unteren Zeile der Tabelle sind zunächst die Summen der Kodierereignisse für die ExpertInnen (481) und die NovizInnen (227) und schließlich die Gesamtsumme (708) aufgeführt. Es ließen sich also 708 Textstellen kodieren, die für die Aktivierung des Wissens der ProbandInnen stehen. Dies ist eine achtbare Anzahl an Nennungen, die zeigt, dass die verwendeten Kategorien ein sinnvolles Instrumentarium zur Kodierung der Interviews darstellen.

**Tab. 38:** Anzahl der vorgenommenen Kodierungen für die inhaltliche und formale Dimension

Faktoren des Konstrukts		ExpertInnen	NovizInnen	Gesamt
<b>Analysekompetenz</b>				
Inhaltliche Dimension	Didaktisch-methodisches Wissen	223	136	<b>359</b>
	Fachliches (incl. fachdidaktisches) Wissen	124	55	<b>179</b>
Formale Dimension	Verstehenstiefe	134	36	<b>170</b>
<b>Summe</b>		<b>481</b>	<b>227</b>	<b>708</b>

Aufgrund der in Tabelle 38 dargestellten Ergebnisse kann zusammenfassend gesagt werden, dass die Hypothesen H1, H2 und H3 bestätigt sind: ExpertInnen aktivierten in deutlich umfangreicherem Maße sowohl didaktisch-methodisches Wissen (H1) als auch fachliches (incl. fachdidaktisches) Wissen (H2) und weisen darüber hinaus mehr Nennungen hinsichtlich der Verstehenstiefe auf (H3).

<sup>82</sup> Die Ergebnisse haben *Vorläufigkeitscharakter*, da erst in Teilkapitel 5.4 (s. S. 156ff.) erörtert wird, ob die vergebenen Kategorien entsprechende Reliabilitätswerte aufweisen und somit in der Tabelle erhalten bleiben.

Während die Werte in Tabelle 38 die Gesamtzahl der Wissensaktivierungen in beiden Dimensionen zeigen, werden in den folgenden drei Tabellen (Tab. 39-41) differenzierte Angaben darüber gemacht, welche *spezifischen* Wissensfacetten die ExpertInnen und NovizInnen hinsichtlich des didaktisch-methodischen Wissens, des fachlichen (incl. fachdidaktischen) Wissens und der Verstehenstiefe aktivierten.

#### – Anzahl der Aktivierungen von Kategorien für didaktisch-methodisches Wissen

In Tabelle 39 sind in der linken Spalte die einzelnen Facetten des didaktisch-methodischen Wissens aufgeführt. In der mittleren Spalte (ExpertInnen) und in der rechten Spalte (NovizInnen) sind die Anzahl der vorgenommenen Kodierungen angegeben, also die Anzahl, die zeigt, wie oft eine entsprechende Wissensfacette aktiviert wurde.

Anhand der Tabelle 39 sieht man, dass einige Wissensfacetten sowohl von den ExpertInnen als auch den NovizInnen häufiger aktiviert wurden. Die drei am häufigsten aktivierten Wissensfacetten sowohl bei den ExpertInnen als auch bei den NovizInnen entsprechen den Kategorien *Lernatmosphäre* (ExpertInnen: 34/NovizInnen: 18), *Diskursivität* (32/30) und *Aktivierung/Motivierung zur geistigen Selbsttätigkeit* (31/25). Umgekehrt gibt es aber auch Wissensfacetten, die von beiden Gruppen weitaus weniger aktiviert wurden. So rekurrierten die ExpertInnen selten auf die Kategorien *Anwendung/Transfer* (11mal), *Lebensweltbezug* (9mal) und *Individualisierung und Differenzierung* (4mal). Die NovizInnen aktivierten die Wissensfacetten *Zielorientierung* (4mal), *Exemplarizität* (3mal) und *Individualisierung/Differenzierung* (0mal) in geringem Maße bzw. gar nicht.

Dass sowohl ExpertInnen als auch NovizInnen einige Wissensfacetten seltener aktivierten könnte mit der Faktur der zu analysierenden Unterrichtsstunde in Zusammenhang stehen. Nicht alle der angeführten Wissensfacetten bestimmen gleichermaßen den realen Verlauf dieser Unterrichtsstunde (s. Kap. 3.1.1, S. 76). So könnte z.B. die geringe Aktivierung der Kategorie *Individualisierung/Differenzierung* (ExpertInnen: 4/NovizInnen: 0) damit zusammenhängen, dass der videografierte Unterricht beinahe keine Möglichkeit für Außenstehende bot, Aussagen über die Homo- bzw. Heterogenität der Lerngruppe zu treffen.

**Tab. 39:** Anzahl der vorgenommenen Kodierungen für das didaktisch-methodische Wissen der ExpertInnen und NovizInnen

Didaktisch-methodisches Wissen	ExpertInnen	NovizInnen
Zielorientierung	25	4
Strukturierung	19	16
Klarheit	15	5
Exemplarizität	17	3
Lebensweltbezug des Themas	9	9
Verknüpfung mit dem Vorwissen	13	6
Aktivierung/Motivierung zur geistigen Selbsttätigkeit	31	25
Diskursivität des Unterrichtsprozesses	32	30
Individualisierung/Differenzierung	4	0
Ergebnissicherung/Erfolgskontrolle	13	11
Anwendung/Transfer	11	9
Lernatmosphäre	34	18
<b>Summe</b>	<b>223</b>	<b>136</b>

Darüber hinaus waren die großen Unterschiede in den Wissensaktivierungen zwischen den ExpertInnen und NovizInnen hinsichtlich der Kategorien *Zielorientierung* (ExpertInnen: 25/NovizInnen: 4), *Klarheit* (15/5), *Exemplarizität* (17/3) und *Vorwissen* (13/6) auffällig. Diese Unterschiede sind zum einen mehr oder weniger erwartbar, denn es ist davon auszugehen, dass die ExpertInnen einen größeren Wissensfundus als die NovizInnen besitzen und daher häufiger ihr didaktisch-methodisches Wissen aktivieren können. Zum anderen könnten diese quantitativen Unterschiede auch auf eine andere Qualität in der Wahrnehmung von ExpertInnen hinweisen. In verschiedenen Studien (s. z.B. Berliner, 2001; König et al., 2014; van den Bogert, van Bruggen, Kostons, & Jochems, 2014; Wolff, van den Bogert, Jarodzka, & Boshuizen, 2015) konnte gezeigt werden, dass ExpertInnen (erfahrene Lehrpersonen) die Fähigkeit besitzen, ihre Aufmerksamkeit sowohl auf den gesamten Lehrprozess (z.B. Zielorientierung) als auch auf die genaue Registrierung wichtiger Details (z.B. Auswahl entsprechender Beispiele) zu lenken, während die NovizInnen diese Fähigkeiten erst noch entwickeln bzw. weiterentwickeln müssen.

– **Anzahl der Aktivierungen von Kategorien für fachliches (incl. fachdidaktisches) Wissen**

In der folgenden Tabelle 40 sind die Aktivierungshäufigkeiten des fachlichen (incl. fachdidaktischen) Wissens der ExpertInnen und NovizInnen zusammengefasst. In der linken Spalte sind wiederum die einzelnen Wissensfacetten angeführt und in der mittleren und rechten Spalte findet sich die Anzahl der Kodierungen für beide Statusgruppen.

**Tab. 40:** Anzahl der vorgenommenen Kodierungen für das fachliche (incl. fachdidaktische) Wissen der ExpertInnen und NovizInnen

Fachliches (incl. fachdidaktisches) Wissen	ExpertInnen	NovizInnen
Physikalische Begriffsbildung	21	10
Durcharbeiten von Fachbegriffen	9	4
Experimentieren als zentrale Fachmethode	19	1
(weitere) fachliche Arbeitsweisen	36	20
Historische Genese physikalischen Wissen	7	4
Produktiver Umgang mit Fehlern	5	5
Lernen als konstruktive Tätigkeit von SchülerInnen	27	11
<b>Summe</b>	<b>124</b>	<b>55</b>

Mit Blick auf diese Tabelle wird deutlich, dass, analog zu den didaktisch-methodischen Wissensfacetten, auch bei den fachlichen (incl. fachdidaktischen) Wissensfacetten auf einzelne häufiger und auf andere in geringerem Maße zurückgegriffen wurde.

Die am häufigsten aktivierten fachlichen (incl. fachdidaktischen) Wissensfacetten sind bei ExpertInnen und NovizInnen *identisch*: *(weitere) fachliche Arbeitsweisen* (ExpertInnen: 36/NovizInnen: 20), *Lernen als konstruktive Tätigkeit von SchülerInnen* (27/11) und *Physikalische Begriffsbildung* (21/10). Im umgekehrten Fall zeigt sich, dass sowohl ExpertInnen als auch NovizInnen ähnliche Wissensfacetten in geringerem Ausmaß aktivierten: *Durcharbeiten von Fachbegriffen* (ExpertInnen: 9/NovizInnen: 4), *Historische Genese physikalischen Wissens* (7/4) und *Produktiver Umgang mit Fehlern* (5/5).

Ein deutlicher Unterschied besteht bei der Aktivierung der Kategorie *Experimentieren als zentrale Fachmethode*. Während ExpertInnen diese Kategorie 19mal aktivierten, griffen NovizInnen auf diese nur einmal zurück. Dieser Unterschied könnte daran liegen, dass NovizInnen (noch) nicht in der Lage sind, die exponierte Stellung des Experiments im Physik-Unterricht zu erkennen, und sie daher dieser wesentlichen Erkenntnisquelle nur wenig Beachtung schenkten (s. Kap. 3.2.3, S. 104).

#### – Anzahl der Aktivierungen von Kategorien für die Verstehenstiefe

In der folgenden Tabelle 41 sind die Kodierungen in der formalen Dimension, die über die Level der Verstehenstiefe operationalisiert wurden, dargestellt.

**Tab. 41:** Anzahl der vorgenommenen Kodierungen für Verstehenstiefe der ExpertInnen und NovizInnen

Verstehenstiefe	ExpertInnen	NovizInnen
Verstehenstiefe – Level 1	65	21
Verstehenstiefe – Level 2a	43	11
Verstehenstiefe – Level 2b	17	4
Verstehenstiefe – Level 3a	3	0
Verstehenstiefe – Level 3b	6	0
<b>Summe</b>	<b>134</b>	<b>36</b>

Die Tabelle zeigt, dass sich die ExpertInnen von den NovizInnen in *quantitativer* wie *qualitativer* Hinsicht deutlich unterscheiden. In quantitativer Hinsicht wurden 134 Nennungen für die ExpertInnen registriert und 36 Nennungen für die NovizInnen (s. auch Tab. 38, S. 137). Dabei zeigt sich, dass die Kategorie „Verstehenstiefe – Level 1“ (VT-1) von beiden Statusgruppen am häufigsten aktiviert wurde (ExpertInnen: 65/NovizInnen: 21 Nennungen).

Mit zunehmendem Schwierigkeitsgrad und höherer Komplexität der Informationsverarbeitung (s. Kap. 1.4.2 & Kap. 3.3) zeigen sich qualitative Unterschiede zwischen den ExpertInnen und NovizInnen. Diese qualitativen Unterschiede bestätigen die Hypothesen H4 und H5: NovizInnen kamen nicht über „Verstehenstiefe – Level 2b“ hinaus (H4), während ExpertInnen das komplette Kategorienspektrum der Verstehenstiefe abdeckten (H5).

### 5.2.2 Unterschiede zwischen ExpertInnen und NovizInnen hinsichtlich des fachlichen Hintergrundes

Mit den bisher dargestellten Unterschieden in den Wissensaktivierungen zwischen ExpertInnen und NovizInnen kam noch nicht der fachliche Hintergrund der ProbandInnen in den Blick. Wie die Arbeiten von Plöger und Scholl (2014a) sowie Plöger, Scholl und Seifert (2015) aber gezeigt haben, hängt die Ausprägung der Analysekompetenz von Lehrpersonen von ihrem *fachlichen* Hintergrund ab (s. Kap. 1.5.2, S. 40). Im folgenden Auswertungsschritt soll deshalb darüber berichtet werden, ob sich dieser Sachverhalt auch in den Ergebnissen meiner Qualitativen Inhaltsanalyse spiegelt.



Tabelle 42 gleicht der oben dargestellten Tabelle 38 (s. S. 137), wird aber um das Kriterium des *fachlichen* Hintergrunds erweitert. Im linken Teil der Tabelle sind wiederum die Faktoren der inhaltlichen Dimension (didaktisch-methodisches Wissen und fachliches Wissen) und der formalen Dimension (Verstehenstiefe) angeführt. In den beiden mittleren Spalten werden die ExpertInnen und NovizInnen hinsichtlich ihres fachlichen Hintergrunds in drei Gruppen unterteilt. Dabei bedeuten die Kürzel Folgendes: Personen, die das Fach Physik unterrichten bzw. studieren werden unter dem Label „Ph“ subsumiert. Personen, die ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten bzw. studieren, werden mit dem Kürzel „NW“ und Personen, die kein naturwissenschaftliches Fach unterrichten bzw. studieren, mit dem Kürzel „KNW“ bezeichnet (s. Kap. 2.3.4, S. 57f.).

**Tab. 42:** Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für die inhaltliche und formale Dimension unter dem Aspekt des fachlichen Hintergrundes

Faktoren des Konstrukts Analysekompetenz		ExpertInnen			NovizInnen			Summe
		Ph	NW	KNW	Ph	NW	KNW	
Inhaltliche Dimension	Didaktisch-methodisches Wissen	69	102	52	49	45	42	359
	Fachliches (incl. fachdidaktisches) Wissen	41	51	32	29	18	8	179
Formale Dimension	Verstehenstiefe	43	44	47	19	11	6	170
<b>Summen</b>		<b>153</b>	<b>197</b>	<b>131</b>	<b>97</b>	<b>74</b>	<b>56</b>	<b>708</b>

Die für die einzelnen Fachgruppen in Tabelle 42 angeführten Häufigkeiten der Kodierungen bestätigen die Hypothesen H6 und H7: ExpertInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten, aktivierten in umfangreicherem Maße fachliches (incl. fachdidaktisches) Wissen als ExpertInnen, die diese Fächer nicht unterrichten (H6). In Entsprechung dazu aktivierten auch NovizInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten, in umfangreicherem Maße fachliches (incl. fachdidaktisches) Wissen als NovizInnen, die diese Fächer nicht unterrichten (H7). Insgesamt zeigt sich somit, dass „Naturwissenschaftler“ (Ph, NW) den „Nicht-Naturwissenschaftlern“ (KNW) in der Analyse der Unterrichtsstunde deutlich überlegen sind.

Unterschiede zeigten sich darüber hinaus auch innerhalb der Gruppe der „Naturwissenschaftler“: Die ExpertInnen, die ein naturwissenschaftliches Fach und/oder Mathematik unterrichten (NW), aktivierten sowohl didaktisch-methodisches Wissen (102 Nennungen) als auch fachliches (incl. fachdidaktisches) Wissen (51 Nennungen) häufiger als die ExpertInnen, die das Fach Physik (69/41) unterrichten.

Dieses Ergebnis ist in dieser Form erwartungswidrig, denn es war eher davon auszugehen, dass sich die ExpertInnen, die das Fach Physik unterrichten, von den ExpertInnen, die ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten, zumindest hin-

sichtlich des fachlichen (incl. fachdidaktischen) Wissens abheben würden. Die Unterschiede in den Aktivierungen des didaktisch-methodischen Wissens, das *generisches* (fächerübergreifendes) Wissen ist, könnten jedoch eine Folge der Fächerkombination der ExpertInnen-NW sein, da einige von diesen auch ein nicht-naturwissenschaftliches Fach (z.B. Sozialkunde) unterrichten und dadurch vermehrt diese Wissensfacetten aktiviert haben könnten.

Im Vergleich dazu gibt es bei den NovizInnen eine recht homogene Verteilung der Aktivierungshäufigkeiten hinsichtlich des didaktisch-methodischen Wissens (49/45/42) und ein (erwartbares) Gefälle bezüglich des fachlichen (incl. fachdidaktischen) Wissens (29/18/8): Diejenigen, die das Fach Physik studieren, aktivierten mit 29 Nennungen fast viermal so oft fachliches (incl. fachdidaktisches) Wissen als NovizInnen, die kein naturwissenschaftliches Fach studieren. Somit kann dieses Ergebnis als Indiz gedeutet werden, dass der fachliche Hintergrund Auswirkungen auf die Aktivierungshäufigkeiten des fachlichen (incl. fachdidaktischen) Wissens und somit auf die Analysekompetenz hat.

In Bezug auf die *formale* Dimension (Verstehenstiefe) zeigten sich bei den ExpertInnen nur geringe Unterschiede in der Häufigkeit der Kategorienvergabe (43/44/47). Bei den NovizInnen besteht ein (erwartetes) Ungleichgewicht: Während hier bei den „PhysikerInnen“ 19 Wissensaktivierungen registriert wurden, konnten bei den NovizInnen, die ein naturwissenschaftliches Fach und/oder Mathematik studieren sowie keines dieser Fächer studieren, nur sechs Wissensaktivierungen gefunden werden, sodass dieses Ergebnis durchaus als Indiz für eine tiefere Verstehensleistung der zu analysierenden Unterrichtsstunde infolge des fachlichen Hintergrundes gedeutet werden kann.

In Analogie zum bisherigen Vorgehen soll nun ein differenzierterer Blick auf die Aktivierung *spezifischer* Wissensfacetten geworfen werden, um auf diese Weise detailliertere Unterschiede zwischen ExpertInnen und NovizInnen hinsichtlich des didaktisch-methodischen Wissens, des fachlichen (incl. fachdidaktischen) Wissens und der Verstehenstiefe in Abhängigkeit vom fachlichen Hintergrund aufzeigen zu können.

#### – Anzahl der Aktivierungen von Kategorien für didaktisch-methodisches Wissen unter dem Aspekt des fachlichen Hintergrundes

In der linken Spalte von Tabelle 43 sind die Kategorien für die Facetten des didaktisch-methodischen Wissens angeführt und in den mittleren Spalten sind die Nennungen der ExpertInnen und NovizInnen hinsichtlich ihres fachlichen Hintergrundes aufgeteilt.

**Tab. 43:** Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für das didaktisch-methodische Wissen unter dem Aspekt der Fachzugehörigkeit

Kategorien für didaktisch-methodisches Wissen	ExpertInnen			NovizInnen			Summe
	Ph	NW	KNW	Ph	NW	KNW	
Zielorientierung	7	15	3	1	1	2	29
Strukturierung	7	11	1	8	4	4	35
Klarheit	5	8	2	2	2	1	20
Exemplarizität	5	10	2	2	0	1	20
Lebensweltbezug des Themas	3	6	0	4	3	2	18
Verknüpfung mit dem Vorwissen	3	6	4	1	2	3	19
Aktivierung/Motivierung zur geistigen Selbsttätigkeit	10	13	8	8	10	7	56
Diskursivität des Unterrichtsprozesses	9	11	12	9	12	9	62
Individualisierung/Differenzierung	1	1	2	0	0	0	4
Ergebnissicherung/Erfolgskontrolle	4	6	3	4	2	5	24
Anwendung/Transfer	3	4	4	4	4	1	20
Lernatmosphäre	12	11	11	6	5	7	52
<b>Summe</b>	<b>69</b>	<b>102</b>	<b>52</b>	<b>49</b>	<b>45</b>	<b>42</b>	<b>359</b>

Mit Blick auf die Tabelle 43 wird deutlich, dass von den ExpertInnen (Ph, NW, KNW) einige Wissensfacetten häufiger und andere in geringerem Maße aktiviert wurden. Zum Beispiel gab es hinsichtlich der Aktivierungen der Facetten *Lernatmosphäre* (ExpertInnen-Ph: 12/ExpertInnen-NW: 11/ExpertInnen-KNW: 11), *Diskursivität des Unterrichtsprozesses* (9/11/12) und *Aktivierung/Motivierung zur geistigen Selbsttätigkeit* (10/13/8) nur geringe Unterschiede.

Deutliche Unterschiede zeigten sich bezüglich der Aktivierungen der Wissensfacetten *Zielorientierung* (7/15/3), *Strukturierung* (7/11/1) und *Exemplarizität* (5/10/2). Vor allem die ExpertInnen-NW aktivierten diese Wissensfacetten in besonderem Maße. Dies könnte, wie bereits weiter oben angesprochen, daran liegen, dass einige dieser ExpertInnen-NW sowohl ein naturwissenschaftliches Fach (z.B. Biologie) als auch ein nicht-naturwissenschaftliches Fach (z.B. Sozialkunde) unterrichten.

In der Gruppe der NovizInnen scheint der fachliche Hintergrund keine Auswirkungen auf die Aktivierungshäufigkeit des didaktisch-methodischen Wissens zu haben, denn sowohl die Gesamtzahl aller Aktivierungen des didaktisch-methodischen Wissens (NovizInnen-Ph: 49/NovizInnen-NW: 45/NovizInnen-KNW: 42) als auch die Aktivierung der einzelnen Wissensfacetten zeigten keine gravierenden Unterschiede zwischen den NovizInnen. Da es sich beim didaktisch-methodischen Wissen um generisches (fächerübergreifendes) Wissen handelt, sind diese geringen Unterschiede erwartbar, da die NovizInnen allesamt

Studierende an der Universität sind und einen vergleichbaren Wissensfundus besitzen dürften.

Interessant scheint darüber hinaus der Vergleich zwischen den ExpertInnen, die weder Physik noch ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten (ExpertInnen-KNW) und den NovizInnen, die das Fach Physik studieren (NovizInnen-Ph), zu sein, denn die gefundenen Unterschiede zwischen diesen beiden Gruppen sind eher gering. Die ExpertInnen-KNW aktivierten ihr didaktisch-methodisches Wissen 52mal und die NovizInnen-Ph 49mal. Die Wissensfacetten *Strukturierung* (ExpertInnen KNW: 1/ NovizInnen-Ph: 8) und *Lebensweltbezug des Themas* (0/4) wurden von den Physik-NovizInnen sogar häufiger aktiviert.

Erklärungen für dieses nicht erwartungskonforme Ergebnis abzugeben ist riskant, weil sie durch diese wenigen Daten kaum gedeckt sind: Möglicherweise gelingt es den hier befragten ExpertInnen-KNW nicht hinreichend, ihr didaktisch-methodisches Wissen auf diese konkrete Unterrichtsstunde anzuwenden, weil ihnen zunächst einmal das notwendige fachliche Wissen zur Beurteilung dieser Unterrichtsstunde fehlt.

– **Anzahl der Aktivierungen von Kategorien für fachliches (incl. fachdidaktisches) Wissen unter dem Aspekt des fachlichen Hintergrundes**

In Tabelle 44 sind die Kategorien für die Facetten des fachlichen (incl. fachdidaktischen) Wissens angeführt und in den folgenden Spalten die Wissensaktivierungen der ExpertInnen und NovizInnen hinsichtlich ihres fachlichen Hintergrundes.

**Tab. 44:** Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für das fachliche (incl. fachdidaktische) Wissen unter dem Aspekt der Fachzugehörigkeit

Kategorien für fachliches (incl. fachdidaktisches) Wissen	ExpertInnen			NovizInnen			Summe
	Ph	NW	KNW	Ph	NW	KNW	
Physikalische Begriffsbildung	5	12	4	7	3	0	31
Durcharbeiten von Fachbegriffen	3	2	4	3	1	0	13
Experimentieren als zentrale Fachmethode	4	10	5	0	1	0	20
(weitere) fachliche Arbeitsweisen	15	14	7	12	5	3	56
Historische Genese physikalischen Wissen	3	3	1	1	2	1	11
Produktiver Umgang mit Fehlern	2	3	0	2	2	1	10
Lernen als konstruktive Tätigkeit von SchülerInnen	9	7	11	4	4	3	38
<b>Summe</b>	<b>41</b>	<b>51</b>	<b>32</b>	<b>29</b>	<b>18</b>	<b>8</b>	<b>179</b>

Anhand der Ergebnisse wird ersichtlich, dass die ExpertInnen-KNW über fast alle Kategorien hinweg weniger Wissensaktivierungen aufwiesen als die Gruppe der „Naturwissenschaftler“ (Ph; NW), sodass dieses Ergebnis die Annahme bestätigt, dass der fachliche Hintergrund bei der Analyse von Unterricht eine zentrale Rolle spielt. Auffällig bei diesen Ergebnissen ist, dass die ExpertInnen-NW im Vergleich zu den Physik-ExpertInnen die Kategorien *Physikalische Begriffsbildung* (ExpertInnen-NW: 12/ExpertInnen-Ph: 5) und *Experimentieren als zentrale Fachmethode* (10/4), deutlich häufiger aktivierten.

Die Ergebnisse der NovizInnen sind erwartungskonform: Die NovizInnen-Ph aktivierten in fast allen Kategorien mehr Wissensfacetten als die anderen NovizInnen. Dabei zeigten sich klare Unterschiede bei den Kategorien *Physikalische Begriffsbildung* und *(weitere) fachliche Arbeitsweisen*, die von den anderen NovizInnen-Gruppen weniger bzw. gar nicht aktiviert wurden. Dieses Ergebnis weist darauf hin, dass der fachliche Hintergrund die Analyseleistung und somit die Aktivierungshäufigkeit spezifischer Wissensfacetten beeinflusst.

Darüber hinaus fallen die Unterschiede zwischen den ExpertInnen-KNW und den NovizInnen-Ph erneut gering aus (ExpertInnen-KNW: 32/NovizInnen-Ph: 29 Wissensaktivierungen). Während die ExpertInnen-KNW vornehmlich die Kategorie *Lernen als konstruktive Tätigkeit von SchülerInnen* aktivierten (ExpertInnen-KNW: 11/NovizInnen-Ph: 4 Wissensaktivierungen), zeigten die NovizInnen-Ph Vorteile bei der Aktivierung der Kategorie *(weitere) fachliche Arbeitsweisen* (ExpertInnen-KNW: 7/NovizInnen-Ph: 12).

#### – Anzahl der Aktivierungen von Kategorien für die Verstehenstiefe hinsichtlich des fachlichen Hintergrunds

Nachdem die Ergebnisse für die inhaltliche Dimension unter dem Aspekt des fachlichen Hintergrunds der ProbandInnen ausgewertet wurden, sollen nun die vorgenommenen Kodierungen für die Kategorien der Verstehenstiefe betrachtet werden. In der linken Spalte von Tabelle 45 sind die Kategorien für die Verstehenstiefe angeführt. In den folgenden Spalten finden sich die Häufigkeiten für die Kategorienvergabe der ExpertInnen und NovizInnen unter dem Gesichtspunkt des fachlichen Hintergrunds.

**Tab. 45:** Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für Verstehenstiefe unter dem Aspekt der Fachzugehörigkeit

Kategorien für Verstehenstiefe	ExpertInnen			NovizInnen			Summe
	Ph	NW	KNW	Ph	NW	KNW	
Verstehenstiefe – Level 1	20	23	22	12	5	4	86
Verstehenstiefe – Level 2a	14	16	13	7	3	1	54
Verstehenstiefe – Level 2b	4	4	9	0	3	1	21
Verstehenstiefe – Level 3a	2	0	1	0	0	0	3
Verstehenstiefe – Level 3b	3	1	2	0	0	0	6
<b>Summe</b>	<b>43</b>	<b>44</b>	<b>47</b>	<b>19</b>	<b>11</b>	<b>6</b>	<b>170</b>

Die ExpertInnen unterscheiden sich hinsichtlich der Gesamtzahl an vorgenommenen Kodierungen für die Verstehenstiefe nur gering voneinander (ExpertInnen-Ph: 43/ExpertInnen-NW: 44/ExpertInnen-KNW: 47 Nennungen). Dabei sind die Kodierhäufigkeiten auf

den unteren Leveln der Verstehenstiefe recht homogen verteilt (z.B. *Verstehenstiefe – Level 1*: 20/23/22). Unterschiede ergaben sich auf den höheren Leveln der Verstehenstiefe: ExpertInnen, die das Fach Physik unterrichten, deckten das *komplette* Kategorienspektrum ab. Sie zeichneten sich also dadurch aus, dass sie nicht nur einzelne didaktisch-methodische Einheiten unter dem Aspekt der Lernwirksamkeit beurteilen konnten, sondern den *gesamten* Unterrichtsverlauf vor diesem Hintergrund zu bewerten vermochten und die zu Beginn der zu analysierenden Unterrichtsstunde antizipierte Beurteilung über die Analyse des *gesamten* Unterrichtsverlaufs hinweg verfolgten. Ein wenig überraschend ist das Ergebnis der ExpertInnen-KNW, denn für diese Gruppe konnten nicht nur die meisten Kodierungen insgesamt vorgenommen werden, sondern diese erbrachten ebenfalls Analyseleistungen bis zum höchsten Level (*Verstehenstiefe – Level 3b*). Daher kann Hypothese H8 nicht bestätigt werden: ExpertInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten, erbringen *keine* tiefere Analyse des Unterrichts (höhere Level der Verstehenstiefe) als ExpertInnen, die diese Fächer nicht unterrichten.

In der Gruppe der NovizInnen zeigen sich deutliche Unterschiede hinsichtlich der Tiefe der erbrachten Analyseleistungen: NovizInnen, die das Fach Physik studieren, kamen auf 19 Nennungen, NovizInnen, die ein anderes naturwissenschaftliches Fach und/oder Mathematik studieren kamen auf 11 Nennungen und für NovizInnen, die weder Physik noch ein anderes naturwissenschaftliches Fach und/oder Mathematik studieren, konnten nur 6 Fundstellen registriert werden. Überwiegend wurden Fundstellen mit *Verstehenstiefe – Level 1* (NovizInnen-Ph: 12 Nennungen/NovizInnen-NW: 5/NovizInnen-KNW: 4) und vereinzelt Fundstellen mit *Verstehenstiefe – Level 2a* kodiert (7/3/1). Damit zeigt sich, dass die NovizInnen nur bedingt Alternativen zu didaktisch-methodischen Einheiten angeben können und noch nicht in der Lage sind, den kompletten Unterrichtsverlauf in seiner Gänze zu erfassen und vor dem Hintergrund der Lernwirksamkeit bewerten zu können. Zusammenfassend können diese Befunde als Bestätigung der Hypothese H9 gedeutet werden: NovizInnen, die das Fach Physik oder ein anderes naturwissenschaftliches Fach und/oder Mathematik studieren, erzielen eine tiefere Analyse des Unterrichts.

### 5.3 Vorstellung beispielhafter Aussagen

Die bisherigen Ergebnisse zeigten numerische Unterschiede in den Wissensaktivierungen zwischen ExpertInnen und NovizInnen. Bevor diese Ergebnisse in Teilkapitel 5.4 für entsprechende Reliabilitätsanalysen und Trennschärferechnungen genutzt werden, soll in diesem Teilkapitel zunächst gezeigt werden, welche konkreten Aussagen sich hinter den im vorangegangenen Teil präsentierten „nackten“ Zahlen verbergen. Dabei werden Aussagen, die von den ExpertInnen stammen mit „E“ eingeleitet und Aussagen der NovizInnen mit „N“ kenntlich gemacht.<sup>83</sup>

---

<sup>83</sup> Die angeführten Textbelege wurden zur besseren Lesbarkeit nachbearbeitet und als *literarische Umschrift* realisiert.

### 5.3.1 Beispielhafte Aussagen für die inhaltliche Dimension

#### 5.3.1.1 Aussagen für didaktisch-methodisches Wissen

##### Zielorientierung

E: „... die Bedingungen, die jetzt ... vorhanden sein müssen, um auf diese Gesetzmäßigkeit zu kommen, die gibt er in gewisser Weise vor. Die kann ich natürlich sehr schlecht erarbeiten, wenn die Zielesetzung nicht klar war. Das ist ein Hin und Her ... Die fehlende Zieltransparenz führt dazu, dass ... die Bedingungen, die Fakten, Parameter nicht erarbeitet werden können mit den Schülern.“

N: „Jetzt sollte da, glaube ich, nach diesem Gesetz gesucht werden. Das fand ich ging sehr schnell und er hat von irgendeinem Protokoll gesprochen, wo ich gar nicht genau wusste, wovon jetzt die Rede war. Er hat dann auch diese Folie aufgelegt, hat dann selber irgendwas erklärt und auch nicht gefragt, ob die Schüler mitkommen oder ob sie überhaupt wissen, wovon jetzt gerade die Rede ist. Also ich zum Beispiel wusste jetzt nicht so genau, ... was er da erklären möchte.“

##### Strukturierung

E: „Also das ist ein entscheidender Punkt, dass die Phasen des Unterrichts eigentlich vollkommen isoliert nebeneinander stehen und dass es deswegen natürlich auch für die Schüler schwer ist, dem roten Faden der Stunde zu folgen.“

N: „Der [Lehrer] springt auch manchmal; der fängt an der einen Stelle an, dann springt er wieder [zu der Frage], warum die Linse kreisförmig ist, dann geht er wieder zum Brechungsgesetz [über]. Das ist ganz verwirrend ... Wenn ich da sitzen würde, ich hätte irgendwie keine Vorstellung.“

##### Klarheit

E: „Der Lehrer ist auch nicht besonders präzise in seiner fachlichen Terminologie. Für ihn ist alles Lichtbrechung, ohne die Begrifflichkeit einmal klar definiert zu haben, was eigentlich eine Brechung bedeutet. Also dass eine Welle oder ein Strahl, je nachdem ob ich jetzt von der geometrischen Optik oder von der Wellenoptik ausgehe, dass eben ein Strahl von einem Medium über eine Phasengrenze in ein anderes Medium, das eventuell andere Eigenschaften wie Dichte hat, übertritt und das dann der Strahl eine Ablenkung erfährt. Also ein sehr diffuses Begriffssystem.“

N: „Also negativ oder was schwierig sein könnte, ist eben wirklich die Fülle an Informationen, die die Schüler bekommen. Das ist wirklich schon sehr viel, weit mehr als man denen zumuten kann.“

##### Exemplarizität

E: [Man müsste] „sich dieser ... Sache ... in einem Dreischritt nähern ... Das Erste ist ja ..., sich dem Beispiel intensiv zu nähern, also ... diese Beispielerhellung, die Klärung dessen, worum es überhaupt geht. Dann käme irgendwann tatsächlich so etwas wie die Abstraktion dieser ganzen Sache, nachdem man sich mit dem Konkreten beschäftigt hat. Da sind durchaus seine Materialien ... funktional, wenn er anders damit umgeht. Und dann



käme aber der Schritt ... der Rekonkretisierung ..., die Rückbindung auf dieses ... Beispiel. Und dann haben die Schüler ja auch andere Beispiele genannt. Das heißt: Ganz spannend wäre zu gucken, wie sieht es denn mit den anderen Beispielen aus ... Er macht es im Grunde genommen ... nicht ... Das heißt, er benutzt ... die Beispiele nicht als unterrichtstragende Beispiele ... Das waren nur Aufhänger für ihn, um ... sehr schnell zu dieser Formel zu kommen, mit der jetzt gerechnet wird. ... In der Tat kann man fragen: ... Wird das Fach damit eigentlich zur Geltung gebracht? ... Und man kann aus der Perspektive der Lerner fragen: Verankert sich da eigentlich irgendetwas im Sinne von Lernzuwachs?

N: „Also das Beispiel mit dem Stab finde ich sehr gut, denn das ist ja ein Phänomen, was auch Kinder selber schon einmal gesehen haben, wenn sie einen Löffel in ein Glas Wasser reingetan haben, dann sehen die, der Löffel ist irgendwie geknickt. Das Beispiel mit der Sonne ist auch okay, aber ich weiß halt nicht, was die vorher schon gemacht haben. Für den Schüler finde ich es selber nicht so ganz ersichtlich, was das jetzt mit Lichtbrechung zu tun hat. Man sieht zwar, da ist was verzerrt, aber da ist dann noch nicht klar, was da an Lichtbrechung passiert ... Ein Laie, der gar nichts damit gemacht hat oder von diesem Begriff noch nichts weiß, kann damit nichts anfangen.“

### **Lebensweltbezug des Themas**

E: „Und jetzt müsste eigentlich von [der] Schülerseite die Frage kommen: Mein Onkel, der hat so eine dicke Brille. Warum kann der im Gestell nicht so eine Folie verwenden? Also, ... Physik ist ja nun wirklich ein Fach, was erlebbare Umwelt, erlebbare Natur, beschreibt. Dass muss man auch nutzen.“

N: „Insgesamt fand ich es relativ langweilig, so an dieses Thema ranzugehen. Ich finde, ... er sollte vielmehr Bezug zur Realität des Schülers schaffen, um das Ganze handfest zu machen. Da wurde relativ wenig verwendet, während besonders bei dem Thema, glaube ich, man hätte mehr verwenden können.“

### **Verknüpfung mit dem Vorwissen**

E: „Man kann ... sagen, dass die Stunde ... ohne Rekurs auf das Vorwissen ... auskommt. Er ruft es ... nicht ab, das wäre eine Möglichkeit, um das Experiment zu konzipieren. ... Häufig lernt man ja nicht im Nachvollzug, ... sondern gerade in der Differenz ... zu bestehendem Wissen. Das nutzt er ... gar nicht. ... Ich finde, ... die Vorkenntnis oder die Vorerfahrung ... spielt keine Rolle.“

N: „Er hat die Schüler mit einbezogen in diesem Einstieg und versucht, auf Vorwissen zurückzugreifen. Er hat dann Bilder gezeigt mit PowerPoint, wo Lichtbrechung zu sehen war und hat dann eine kleinere Erklärung abgegeben und versucht dann den Umschwung, ein selbstgestaltetes Experiment, um diese Lichtbrechung neu zu entdecken. ... Er lässt die Schüler jetzt aber nicht selber auf dieses Thema kommen, sondern er führt schon eigenständig ein. Er bezieht sich dann aufs Vorwissen [der Schüler] und fragt, ‚Wo gibt es so etwas?‘ und ‚Kennt ihr das?‘, aber er interessiert sich nicht dafür.“

### **Aktivierung/Motivierung zur geistigen Selbsttätigkeit**

E: „Ihm gelingt es überhaupt nicht, die Aufmerksamkeit der Schüler zu fokussieren. ... Er hätte das Tafelbild ja auch mit den Schülern entwickeln können. ... So geht das an den Schülern vorbei. Es war ja auch auffällig, das Konzentrationsvermögen, das war weg. Die Schüler haben [sich] teilweise gelangweilt die Sache angeschaut.“

N: „Das geht ein bisschen in die Richtung, dass wir [später] als Lehrer nicht versuchen, ... die Erkenntnis in die Schülerköpfe hereinzudrücken, sondern versuchen, dass sie es schaffen, auch auf eigene Ideen zu kommen, auf eigene Ergebnisse. Auch wenn die falsch sind. Aber ein eigenes Ergebnis ist immerhin mehr wert als ein Ergebnis, was man sozusagen wie auf einem Präsentierteller serviert bekommt.“

### **Diskursivität des Unterrichtsprozesses/Reflexion von Inhalt und Methode**

E: „Lernen heißt für die Schüler und auch für den Lehrer die richtigen Werte finden. Seien es die Werte des Lehrers oder die von Snellius, dass das Richtige rauskommt. Diese Form des Unterrichts führt langfristig dazu, dass die Schüler immer darauf erpicht sind, möglichst genau das zu finden, was der Lehrer zu Hause vorgedacht hat und damit sind sie nicht bei sich, sondern sie sind immer bei dem Lehrer und denken, was will er jetzt als nächstes haben. Und das kann nicht Aufgabe eines guten Unterrichts sein.“

N: „Negativ finde ich auch, dass er viel zu wenig die Schüler miteingebunden hat ... viel zu wenig mit denen gearbeitet hat, sondern das war halt ein reiner Lehrervortrag. Es gibt natürlich Stunden, wo es unumgänglich ist, mal Lehrervorträge zu halten, weil das Thema ... nicht die Möglichkeit gibt, dass man die Schüler komplett miteinbaut. ... Aber das ist ein Thema, was man sehr sehr gut mit den Schülern machen kann, wo auch vor allen Dingen die Schüler selber das lernen müssen, wo der Lehrer sich nicht großartig mit einbringen muss.“

### **Individualisierung/Differenzierung**

E: [Der Lehrer] „verschenkt unglaublich viele Chancen. Vor allen Dingen, wenn es darum geht, Schülern eigenständige Lernwege zu eröffnen, sie nachdenklich zu machen, wenn in der Weise alles von vorne sozusagen vorgekaut, vorgedacht wird ... Aufgrund der Tatsache, dass Schüler mit unterschiedlichem Vorwissen, mit unterschiedlichem Interesse, auch unterschiedlich motiviert an denselben Gegenstand herangehen, spricht eigentlich viel mehr dafür, dass jeder in seinem eigenen Lerntempo auch die Chance bekommt, selbstständig ... selbsttätig Dinge durch Experimentieren herauszufinden und sich dadurch auch selber zu bestätigen in ihren eigenen Fähigkeiten.“

N: An dieser Stelle kann keine Aussage der NovizInnen präsentiert werden, da diese die Wissensfacette Individualisierung/Differenzierung nicht aktivierten (s. Tab. 39, S. 138).

### **Ergebnissicherung/Erfolgskontrolle**

E: [Der Lehrer] „selbst weiß überhaupt nicht, was er da gesichert hat bei den Schülern, was die verstanden haben. Die hätten ja mal schauen können, was sollen wir machen. Man hätte einen Schüler dann zusammenfassen lassen können. Er sichert sich also nirgendwo ab, wie das ankommt. Das heißt also, er sieht auch nicht die unterschiedliche Lern- und Leistungsfähigkeit seiner Schüler.“

N: „Die Schüler haben eine ziemlich passive Rolle. Und was irgendwie auch auffällt: Man merkt gar nicht, ob der Lehrer sich wirklich dafür interessiert, ob die Schüler das verstanden haben oder so. Der erzählt halt seine Sachen und findet das alles wahrscheinlich für sich selber ganz logisch. ... Also zumindest ist mir nicht aufgefallen, dass er einmal Rückfragen stellt: ‚So gibt es Fragen?‘ Oder: ‚Habt ihr das verstanden?‘“

### **Anwendung/Transfer**

E: „... der letzte Schritt ... sicher werden [die Schüler] üben, also Kompetenzen dann sichern. Das hat er mit dieser Aufgabe und mit dieser Anwendung versucht, aber die ist dann schon ein Schritt zu weit, und das können Schüler an der Stelle nicht nachvollziehen. Auffällig ist, dass er etwas entspannter wurde, als er mit der Fresnel-Linse experimentiert hat, ... aber das Lernen selbst ... wirkt sehr gedämpft ..., schade.“

N: „Die Linse würde ich nächste Stunde einbauen. Die würde ich nicht am Ende Stunde bringen, sondern ich würde dann wirklich nur das Brechungsgesetz in dieser Stunde behandeln und würde das mit den Linsen am Ende gar nicht mehr machen und auch nicht den Begriff der Linse einbringen, damit die Schüler so eine Art Cut haben und wissen, das ist das Brechungsgesetz und nächste Stunde kommen Linsen.“

### **Lernatmosphäre**

E: „Wenn ich dann noch einmal auf den Umgang Lehrer-Schüler eingehe ... Ich denke, die haben nicht den Eindruck, dass sie als Schüler wirklich ernst genommen werden mit ihrem Wissensstand. Das Alltagswissen wird belächelt. ... Das können wir eigentlich nicht gebrauchen, [denn] die Wissenschaft ist eine ganz andere Ebene. Das heißt also: Als Mensch durchaus sympathisch, aber die Schüler fühlen sich mit Sicherheit nicht kompetent, was das Fach angeht. Die Sitzordnung tut ihr Übriges dazu, das ist ja oft in naturwissenschaftlichen Fachräumen so.“

N: „Ich habe persönlich das Gefühl, dass er zwar die Schüler fragt, aber nicht genug ernst nimmt. Er fragt sie irgendwas, und dann macht er einen Witz über sie. Witze finde ich prinzipiell gut, aber nicht über die Schüler ... Ich finde auch, das mit dem Fisch, ... das ist eigentlich ein super Beispiel gewesen, [da] hätte er mehr loben können.“

## **5.3.1.2 Aussagen für fachliches (incl. fachdidaktisches) Wissen**

### **Physikalische Begriffsbildung**

E: [Die Schüler] „können keine Lösungsstrategien entwickeln. Das hängt auch mit seiner begrifflichen Unexaktheit zusammen ... von der ich glaube, dass er selbst nicht über ein so profundes Fachwissen verfügt, denn jemand, der wirklich das internalisiert hat, der geht mit den Begrifflichkeiten viel klarer und sauberer um und fordert die auch in der Klarheit von den Schülern ein.“

N: „... was ich auch nicht gut fand, er hat immer vom Brechungsgesetz gesprochen, aber ... ich wusste nicht genau, worauf er hinauswill. Was will man vom Brechungsgesetz? Man will irgendwelche Winkel ausrechnen, da hätte das Wort Winkel zum Beispiel mal fallen können. Dann hätte man gewusst, dass man Winkel messen will, dann hätten die Schüler vielleicht schon eher gewusst, was man für einen Versuchsaufbau machen kann.“

Aber wenn man nur Brechungsgesetze [hört] und nicht weiß, worum es geht, dann kann man sich schlecht einen Versuchsaufbau ausdenken. ... Also das war ... völlig fernliegend, da hätte man nie darauf kommen können.“

### **Durcharbeiten von Fachbegriffen**

E: „... ein viel zu schneller Sprung von der Linse zu der Fresnel-Linse. Das war für die Schüler schon nicht mehr nachvollziehbar. ... Er hätte hier den Lernzugewinn definieren [müssen], ... die Schüler müssen erkennen, was wissen wir jetzt mehr im Vergleich zum Anfang, und der letzte Schritt wäre sicher werden und üben, also Kompetenzen dann sichern.“

N: „Also ich frage mich, wie er es geschafft hat, jetzt in die Stunde noch die Linse reinzupressen. Also das muss unglaublich schnell gewesen sein. Ich glaube nicht, dass da noch jemand mitgekommen ist. ... Es wundert mich ein bisschen, dass er direkt auf die Fresnel-Linse kommt, denn normalerweise müsse man sich vorher über Linsen unterhalten, damit man mit Linsen ... klar kommt, sich vorstellen kann, was das ist, was die machen, wie das Brechungsgesetz bei Linsen funktioniert. ... Und jetzt direkt in dieser Stunde, in der er gerade einmal das Brechungsgesetz eingeführt hat, dann direkt auf die Fresnel-Linse zu kommen, das ist zu schnell.“

### **Experimentieren als zentrale Fachmethode**

E: „Der naturwissenschaftliche Erkenntnisweg hätte ja von den Schülern [selbst] überlegt werden können: Wie gehen wir denn jetzt mit unserem Phänomen [um], wir wollen es ja als Physiker ergründen, wir wollen eine Gesetzmäßigkeit aufdecken. Wie machen wir das? – Er nimmt sie nicht mit auf diesen naturwissenschaftlichen Erkenntnisweg, er lässt alles nur nachvollziehen. [Und dann die] nächste Chance, die er nicht genutzt hat: Wie werden wir unser Experiment denn jetzt aufbauen? Die Schüler versuchen, eine Idee einzubringen, auch die wird im Grunde wieder sehr schnell ad acta gelegt. Das heißt also: Die Förderung der Selbstständigkeit beim naturwissenschaftlichen Denken hat er in dieser Sequenz überhaupt nicht berücksichtigt, das sehe ich als ganz extremes Manko in dieser Stunde.“

N: „Die Vorgehensweise müsste praxisorientierter sein. Ich würde eine andere Methodik anwenden, weil solche Experimente, egal ob in der Physik, Chemie oder irgendeiner anderen Naturwissenschaft, sind immer einprägsamer, wenn die Schüler das selber machen und wenn die die Ergebnisse selbst herausfinden und das nicht so passiert, wie er das macht, dass er die Ergebnisse vorne zusammenfasst. Das ist was, was die Schüler sich nicht merken. Also daran erinnert sich kein Schüler später.“

### **(weitere) fachliche Arbeitsweisen**

E: „... es muss am Anfang so einer Stunde, wo es um die Einführung eines neuen Themas geht, da muss ein Phänomen stehen. Die Schüler müssen zunächst mal etwas haben, mit dem sie arbeiten können und das passiert hier gar nicht. Also das Phänomen kommt erst später ... Er sagt, ich habe hier ein Bild, da ist ein Eimer zu sehen, da ist ein Strohhalmin drin, der sieht aus, als wäre der geknickt, ... ist er aber nicht. Damit ist doch sozusagen das Feuer direkt wieder gelöscht, was er versucht anzuzünden.“

N: „Er versucht auf die Formel vom Brechungsgesetz zu kommen und dafür hat er jetzt eine Skizze verwendet, die er an der Tafel vorbereitet hat und dort dann einen Strahlengang angezeichnet hat, wie der dann gebrochen werden würde, und dann hat er nachgefragt, ob es etwas Proportionales gibt, was man aus den Messergebnissen ersehen kann und da hat er nichts gefunden. Also bei mir als Physiker hätten die Messwerte erste Priorität, denn man kann Messfehler machen, das muss man miteinbeziehen. ... Zum Beispiel haben die ersten zwei, drei Werte gestimmt und waren 1,5fach größer, aber die anderen haben dann nicht mehr ins Schema gepasst und werden direkt abgelehnt. Aber das wäre jetzt die Frage, warum passen die ersten Werte und dann nicht mehr. Das könnte doch vielleicht ein Messfehler sein und das müsste man nochmal prüfen.“

### **Historische Genese physikalischen Wissens**

E: „... also was klar erkennbar ist, dass der Lehrer zunächst das Thema rahmt, den Schülern einen Ausblick gibt, auf das, was auf sie zukommt, dass er das Thema historisch einordnet, in einen historischen Kontext einbindet, und sein Bemühen erkennen lässt, hier einen genetischen Weg, [eine] genetische Erkenntnisgewinnung im Schnelldurchgang nachzuvollziehen. Das greift er an späterer Stelle auch noch einmal auf, das ist so ein Grundprinzip, was ich seinem Unterricht entnehme.“

N: „Zunächst fand ich es ganz gut, dass der Lehrer von einem aktuellen Problem erstmal auf die Geschichte des Problems hinweist, also vor vierhundert Jahren war das so und so ... Das wird, soweit ich weiß, im Physikunterricht ganz gerne gemacht, weil es immer eine Entwicklung gab in den verschiedenen Entdeckungen und von daher ist es ganz gut, in der Geschichte anzufangen.“

### **Produktiver Umgang mit Fehlern**

E: „Er versucht halt eben jetzt die Bedingungen mit den Schülern herauszuarbeiten, die gegeben sein müssen. ‚Eine ebene Fläche vielleicht‘, [eine] Vorgabe von ihm, im Grunde suggestiv, ohne dass diese Idee von den Schülern kommt. [Ein] Glaskörper vielleicht, ja weil er durchsichtig ist, und dann sagte eine Schülerin, dass Glas Fasern hat und die können dann ja noch mal brechen. Und dann sagt er einfach nur nein. Er geht nachher auch gar nicht mehr auf diese Aussage der Schülerin ein. Man müsste hier ja doch eine differenzierte Rückmeldung geben, warum das jetzt hier so nicht zutrifft.“

N: „... was mir gleich am Anfang aufgefallen ist, dass eine Schülerin gesagt hat, wie sie sich das vorstellt mit den Glasfasern, dass das Licht an den Fasern irgendwie dann gebrochen wird. Aber da ist er ganz schnell einfach drüber weggegangen. ... doch, mit Fehlvorstellungen von Schülern muss man ... sehr aufpassen, ... man darf die nicht so einfach ignorieren und mit dem überwalzen, was man dann als richtig erachtet, sondern man muss eben diese wirklich in das Unterrichtsgeschehen einbauen und sich vorher Gedanken machen, welche Vorstellungen könnten da schon existieren und die dann benutzen und wirklich am Schluss nochmal ausräumen. ... Man müsste diese Fehlvorstellungen der Schülerin nochmal aufgreifen, um dann zu sagen, ... warum das anders sein müsste.“

## Lernen als konstruktive Tätigkeit von SchülerInnen

E: „... ich kenne das Arbeitsblatt nicht, aber ich vermute, dass es so ähnlich aufgebaut ist, wie das auf der Folie. Die Schüler sollen lediglich die Wörter des Lehrers nachvollziehen, vielleicht sogar aufschreiben. Ob sie überhaupt damit hantieren, wird überhaupt nicht deutlich. Was mich auch erschreckt, ist, dass er präzise sagt, [der Lichtstrahl] muss in die Mitte des Halbkreises. Das ist physikalisch richtig, aber er blockiert alle Denkbewegungen der Schüler.“

N: „Er ist ... nicht nur ein Anleiter, sondern er ist ... der, der das Wissen vermittelt. Jetzt nicht im Sinne von einem Moderator, ... sondern er ist wirklich der Inbegriff des Wissensvermittlers und ich glaube, ... dass er darin seine zentrale Aufgabe sieht, das Wissen zu vermitteln.“

### 5.3.2 Beispielhafte Aussagen für die formale Dimension

#### Aussagen für Verstehenstiefe – Level 1 (VT-1)

(Definition der Kategorie: „Verstehenstiefe – Level 1“ steht für die Fähigkeit, Verbindungen zwischen Situationen oder Handlungen und den dadurch hervorgerufenen Effekten herzustellen.)<sup>84</sup>

E: „Zur Schüleraktivität ist ... alles schnell gesagt, die reduziert sich fast auf Null, auf einige wenige Beiträge. ... Überhaupt [ist] der Unterricht ..., und das liegt an dem zugrunde liegenden Skript, nicht auf Schüleraktivität hin ausgerichtet. Er verschenkt hier Chancen, ganz gravierende Chancen des Physikunterrichts. ... Dass hier keine Schüleraktivität zustande kommt, liegt daran, dass das Unterrichtsskript nicht auf Schüleraktivität hin ausgerichtet ist.“

N: „... und auch diese Schülervorstellung ... ein Mädchen hat ziemlich genau gesagt, wie sie sich das vorstellt mit dem Glas und den Fasern, wo er dann direkt gesagt hat ‚Nee, ich glaube, so sieht das nicht aus‘. Das muss man als Lehrer aufnehmen und damit weiterarbeiten und nicht einfach sagen, dass das nicht so ist, sondern sagen, das ist so und so und einem das versuchen zu erklären, dass man diese Vorstellung aus den Köpfen bekommt, weil ich finde, sonst kann man da keinen vernünftigen Unterricht darauf aufbauen.“

---

<sup>84</sup> Für die Kategorien der Verstehenstiefe werden die Definitionen aus Kapitel 3.3.3 (s. S. 126ff.) mit angegeben, da diese Kategorien nicht selbsterklärend sind.

### Aussagen für Verstehenstiefe – Level 2a (VT-2a)

(Definition der Kategorie: „Verstehenstiefe – Level 2a“ bezieht sich auf die Fähigkeit, begründbare Alternativen zu didaktisch-methodischen Sinneinheiten angeben zu können.)

E: „Das schafft Enttäuschungen ... bei den Schülern, [so] dass sie das Gefühl haben, wir wissen nicht genug, wir haben viel vergessen. Und sie schämen sich. Das schafft merkwürdige Situationen. Aus unserer Sicht gilt es, die von vornherein durch professionelles Handeln zu vermeiden, indem ich Klarheit schaffe. Und das ließe sich leicht so arrangieren. Wir brauchen ... gleich noch einmal Vorwissen über optische Phänomene, hier [über] Brechung. ... Ihr habt jetzt ... drei Minuten Zeit, um euch in Kleingruppen, mit Nachbarn ... zu informieren, und dann spielen wir das, was noch an Vorwissen da ist, ... noch einmal ins Plenum hinein. Das schafft eine saubere Basis für die Weiterarbeit und ist den Schülern gegenüber ehrlich, fair. ... Und die Schüler wissen, wir brauchen Vorwissen, und der Lehrer arbeitet mit dem, was da ist. ... Er macht es sich hier unnötig schwer.“

N: „Allgemein ... geht es immer ums Experimentieren im Physikunterricht und wenn man ein Experiment durchführt, hat man eben die Möglichkeit, alle Schüler sozusagen dran zu beteiligen. ... Da gibt es halt Möglichkeiten, die einfach mal ausprobieren zu lassen oder das, was er jetzt anhand der Bilder an die Wand geworfen hat, einfach mal zu demonstrieren. Er hat auf dem Bild einen Eimer gezeigt, wo ein Stab drin war und da kann man einfach einen Eimer mitbringen, mit Wasser füllen und den Stab reinhalten, ... sodass alle Schüler wirklich dastehen und sehen, ‚aha, so ist es in der Wirklichkeit‘. ... Wenn alle nach vorne müssen, sich um den Eimer stellen, dann sind auch alle dabei und vielleicht [könnte er] auch die Schüler direkter ansprechen und mehr zu Wort kommen lassen. Also einfach mehr Fragen formulieren lassen, sodass jeder Schüler mal kurz was dazu sagen kann und dadurch auch ein höheres Wechseltempo zwischen Lehrer- und Schülersprache entsteht.“

### Aussagen für Verstehenstiefe – Level 2b (VT-2b)

(Definition der Kategorie: „Verstehenstiefe – Level 2b“ steht für die Fähigkeit, Verbindungen zwischen Situationen oder Handlungen herzustellen und die dadurch konstruierten Zusammenhänge durch [hypothetische] Vorgriffe [Antizipationen] oder durch Rückverweise explizieren zu können.)

E: „Was wir eben schon mal gesagt haben! Es ist eigentlich keine wirkliche Problemstellung für den Schüler erkennbar. Und dann die Ausführung, die Auswertung des Experiments erfolgt ja auch nicht, indem er die naturwissenschaftlichen Denkweisen des Schülers aktiviert, sondern er macht es auf seine oberflächliche Art und Weise.“

N: „Ich weiß jetzt nicht, was ich ... dazu sagen soll. Also man hat jetzt gehört, dass es eine Partnerarbeit zwischendurch gab. Das finde ich gut. Jetzt geht es an der Tafel weiter, das finde ich auch gut. Das habe ich ja eben noch gesagt, dass mir das fehlt, dass etwas an der Tafel festgehalten wird. Mal sehen, wie es jetzt weitergeht. Also wenn er dann noch vielleicht ein paar Regeln oder Gesetze an der Tafel auch festhalten würde, dann fände ich das hilfreich.“



### Aussagen für Verstehenstiefe – Level 3a (VT-3a)

(Definition der Kategorie: „Verstehenstiefe – Level 3a“ bezieht sich auf die Fähigkeit, den gesamten Unterrichtsprozess zu erfassen.)

E: „Jetzt gehe ich ein auf das Unterrichtsskript ein ... Es ist ein Kennzeichen seines Unterrichts, ...dass er ... das Problem aufreißt und ... von den Schülern verlangt, dass sie diesen Erkenntnisweg selbständig finden unter seiner Mitwirkung. Ich ... erweitere das Konzept [nun] so, dass es erfolgreicher wird. Es geht zunächst einmal ... in einem ersten Schritt [darum], dass Schüler eine Problemstellung entdecken ... Zweiter Schritt: Die Schüler müssen dazu Vorstellungen entwickeln; es ist meine Pflicht zuzuhören, welche Vorstellungen existieren zu dieser Problemstellung in der Klasse ..., so dass Schüler [sich] angstfrei ohne meine Intervention äußern können. Ich ... sehe ... [dann]: Solche Vorstellungen gibt es, solche Ideen gibt es, da werden Hypothesen gebildet, da werden Erklärungsversuche gemacht, ein ganzes Konglomerat an Ideen, Vorstellungen, ... Hypothesen; die muss ich aufnehmen. ... Also kommen wir dann in die dritte Phase. Ich muss einen Informationsinput geben, mit dem dann die Schüler Erkenntnisse gewinnen, selbstständig, und ein Lernprodukt erstellen ... Der dritte Schritt ist der zentrale: ... Ich gebe Informationen hinein; das kann geschehen über Lernmaterialien, das kann geschehen über einen Lehrervortrag, über ein Experiment, über Arbeitsblätter, wie auch immer. ... Die Schüler ... stellen ausgewählte Lernprodukte im Plenum vor, die dann zur Diskussion gestellt werden, um so [das] Lernen abzugleichen, Ideen in den Raum hinein zu bringen. ... Das schafft automatisch Schüleraktivität und da brauche ich auch nicht nachzufragen. ... Kurzum, dieser Lehrer macht sich das Leben an vielen [Stellen] unendlich schwer, [er] überfordert die Schüler, [und hat dadurch] mäßigen Erfolg.“

N: Hier kann keine Aussage der NovizInnen vorgestellt werden, da diese nicht über Verstehenstiefe – Level 2b hinaus gekommen sind (s. Tab. 41, S. 140).

### Aussagen für Verstehenstiefe – Level 3b (VT-3b)

(Definition der Kategorie: Die Kategorie „Verstehenstiefe – Level 3b“ steht für die fortlaufende Überprüfung des zu Beginn der Analyse hypothetisch unterstellten Unterrichtsverlaufs.)

Dazu möchte ich folgende Anmerkung vorwegschicken: Die folgenden Aussagen beziehen sich auf vier Zeitpunkte, an denen die Videodarbietung unterbrochen und die Person zu einer Stellungnahme aufgefordert wurde. Bereits nach der ersten Sequenz wagt sie eine *Prognose* über den weiteren Unterrichtsverlauf, obwohl sie ihn noch nicht kennt. In den folgenden Aussagen greift die Person dann immer wieder auf die vorangegangenen Sequenzen zurück. Das zeigt, dass sie den *gesamten* Unterrichtsverlauf als *Einheit* verstanden hat. Dieser Level der Verstehenstiefe entspricht der Stufe 5 (VT-3b) des in Abb. 4 dargestellten Modells (s. S. 36).

Im Anschluss an die erste Sequenz: E: „Es ist der Versuch, die Frage der Lichtbrechung mit den Schülern gemeinsam als Problemstellung zu entwickeln ... Man kann allerdings jetzt schon sehen, dass er dieses Prinzip ... nicht durchhalten kann.“

Im Anschluss an die zweite Sequenz: E: „Im Grunde geht es weiter so. ... Normal wäre [es], dass er jetzt mal fragt, welche Ergebnisse habt ihr eigentlich herausbekommen, dass

er die Chance nutzt, Schüler miteinander ins Gespräch kommen zu lassen, dass sie ihre Werte austauschen ... Er nimmt wieder nicht ernst, was die Schüler beizutragen haben.“

Im Anschluss an die dritte Sequenz: E: „Jetzt versucht er wieder die Schüler mit ins Boot zu holen, aber es ist eigentlich zu spät.“

Im Anschluss an die vierte Sequenz: E: „[Das] war noch einmal eine ganz interessante Sequenz, weil sich bestätigt, was sich schon am Anfang andeutete. Er versteht auf der einen Seite durch seine humorvolle Art, die Schüler immer wieder persönlich anzusprechen, aber gleichzeitig macht er hier genau dasselbe, [was] er die ganze Zeit schon gemacht hat.“

N: Da die NovizInnen nicht über Verstehenstiefe – Level 2b hinauskamen, kann hier keine Aussage präsentiert werden (s. Tab. 41, S. 140).

## 5.4 Quantifizierung der Ergebnisse

Die bisher präsentierten Ergebnisse dokumentieren die Häufigkeiten der von Gruppen (z.B. ExpertInnen vs. NovizInnen) und Subgruppen (z.B. ExpertInnen-Physik; NovizInnen-Physik) aktivierten Wissensfacetten. Auf diese Weise ließen sich entsprechende Unterschiede deutlich machen, wie sie sich z.B. in der grundsätzlichen Überlegenheit bei der Wissensaktivierung des didaktisch-methodischen Wissens der ExpertInnen zeigten. Mithilfe der bisherigen Ergebnisse ließen sich auch eine Reihe von Hypothesen prüfen, die die faktorielle Validität des Konstrukts Analysekompetenz betrafen. Insgesamt gesehen, kann daher festgehalten werden, dass sich die bei Erhebung und Interpretation der Daten genutzte Methode der Qualitativen Inhaltsanalyse als hilfreiches Instrument erwiesen hat. Darüber hinaus muss man aber auch die Grenzen dieses methodischen Zugriffs sehen, die sich insbesondere in dreifacher Hinsicht zeigen:

- 1) Es bleibt ungeklärt, ob es sich bei den augenscheinlich deutlichen Unterschieden zwischen ExpertInnen und NovizInnen um statistisch bedeutsame, also signifikante Unterschiede handelt.
- 2) Aus der bisherigen Art der Ergebnisdarstellung wird nicht ersichtlich, ob *einzelne* Personen die gesamte Palette der Facetten (definiert durch die entsprechenden Kategorien) nutzten oder vorwiegend bestimmte Wissensfacetten aktivierten und andere Facetten möglicherweise seltener oder gar nicht. Wenn sich zeigen ließe, dass die einzelnen Personen nicht nur einige, sondern im Prinzip alle Facetten des jeweiligen Faktors der inhaltlichen Dimension (didaktisch-methodisches Wissen oder fachliches Wissen) genutzt haben, um die Unterrichtsstunde zu analysieren, dann wäre dies ein Beleg für die inhaltliche Kohärenz der Facetten *innerhalb eines Faktors*. Darin würde ich dann auch eine nachträgliche Rechtfertigung der Wahl der Kategorien für den jeweiligen Faktor sehen, deren Legitimation ich bisher aus den Forschungen über Prinzipien effektiven Lehrens, aus fachdidaktischen Argumentationen und Ergebnissen der Expertiseforschung hergeleitet hatte (s. Kap. 1 & Kap. 3).
- 3) Die Anzahl der Kodierungen für didaktisch-methodisches Wissen einerseits und für fachliches Wissen andererseits lässt auch keine Schlüsse darüber zu, ob Personen, die

häufiger (seltener) Facetten der einen Wissensart (z.B. didaktisch-methodisches Wissen) aktivierten, auch häufiger (seltener) Facetten der anderen Wissensart (z.B. fachliches Wissen) nutzten. Sollte dies der Fall sein, so wäre dies ein Indiz für die Kohärenz zwischen den beiden Faktoren der inhaltlichen Dimension und somit auch dafür, dass sich beide Wissensarten bei der Analyse von Unterricht effektiv ergänzen.

Antworten auf diese drei Fragen setzen voraus, dass man die qualitativ erhobenen Daten quantitativ weiterverarbeitet: Jede Aussage der einzelnen ProbandInnen repräsentiert – in Analogie zu einer richtigen Lösung eines Items in einem Test – einen „Treffer“ und kann somit als ein erreichter „Punkt“ gezählt werden. Auf diese Weise stellen die Kodierungen quantitative Ausprägungen der Kategorien dar. Unterschiedliche Ausprägungen der einzelnen Variablen ergeben sich aus der Häufigkeit, wie oft eine Person Wissen aktiviert, das einer bestimmten Kategorie zugeordnet werden konnte.

Zur Durchführung entsprechender statistischer Berechnungen wurde in Kapitel 4 eine Sprachregelung eingeführt, die sich an den Begrifflichkeiten für Reliabilitätsanalysen orientiert (s. Kap. 4, S. 133f.). Mit Blick auf die dort eingeführte Regelung werden die verwendeten Kategorien als *Variablen* bezeichnet, die zusammen eine homogene Teil-Skala [Faktor] des Konstrukts Analysekompetenz repräsentieren: In der inhaltlichen Dimension umfasst der Faktor des didaktisch-methodischen Wissens zwölf Variablen (= Kategorien) und der Faktor des fachlichen (incl. fachdidaktischen) Wissens sieben Variablen. In Entsprechung dazu kann dann die inhaltliche Dimension des Kategoriensystems insgesamt als eine Skala mit 19 Variablen interpretiert werden. In der formalen Dimension beinhaltet der Faktor der Verstehenstiefe (unterschiedliche Komplexitätsgrade der Informationsverarbeitung) fünf Variablen.

Auf dieser Basis können die quantitativen Daten dann mit den üblichen statistischen Verfahren<sup>85</sup> ausgewertet und im Hinblick auf die drei offenen Fragen hin untersucht werden:

- 1) Die Unterschiede in den Häufigkeiten der Wissensaktivierungen von ExpertInnen und NovizInnen können auf statistische Signifikanz geprüft werden.
- 2) Die Berechnung der Trennschärfen steht für die Korrelationen (= Stärke des Zusammenhangs) zwischen der Anzahl der Aktivierungen jeder einzelnen Variable (Wissensfacette) und der Anzahl der Aktivierungen für alle anderen Variablen des jeweiligen Faktors. Je höher die Trennschärfe einer Variable (Kategorie) ist, desto besser repräsentiert sie die gesamte Skala des einzelnen Faktors. Zusammen mit den Trennschärfen

---

85 Die Wahl eines geeigneten statistischen Prüfverfahrens hängt im Wesentlichen von zwei Faktoren ab, einerseits vom Skalenniveau der ermittelten Werte und andererseits von der Größe der Stichprobe:

1) Hinsichtlich des Skalenniveaus stellen die in Tabelle 38 (s. S. 137) angeführten Werte intervallskalierte Daten dar: Die einzelnen Kodierereignisse repräsentieren die richtige, d.h. von den KodiererInnen akzeptierte Anwendung von Wissen. Differenzen, die durch unterschiedliche Zahlen bzw. Summen dargestellt werden, können somit als entsprechende Differenzen hinsichtlich der Ausprägung der Wissensaktivierung interpretiert werden.

2) Die in Tabelle 38 angeführten Daten sind allerdings aufgrund einer relativ kleinen Stichprobe mit insgesamt 18 ExpertInnen und NovizInnen zustande gekommen. Bei derart kleinen Stichproben kann zunächst nicht davon ausgegangen werden, dass die Werte einer Normalverteilung folgen, sodass es fraglich ist, ob zur Auswertung der Daten parametrische Prüfverfahren eingesetzt werden könnten (s. z.B. Bortz & Lienert, 2008; Bortz, Lienert, & Boehnke, 2008). Die Diskussion dieser Frage wird weiter unten wieder aufgenommen (s. S. 171).

kann dann auch die Reliabilität der Skalen berechnet werden. Hohe Reliabilitätswerte zeigen an, dass die Kodierungen zu „messgenauen“ Ergebnissen geführt haben.

- 3) Die Stärke der Beziehungen zwischen den Facetten des didaktisch-methodischen und fachlichen Wissens kann durch die Berechnung der Korrelation zwischen der Anzahl der Aktivierungen für das didaktisch-methodische und das fachliche Wissen berechnet werden. Hohe Korrelationen würden indizieren, dass die ProbandInnen beide Wissensarten für eine angemessene Analyse von Unterricht nutzen. Darüber hinaus würde eine hohe Korrelation zusätzliche Evidenz dafür erbringen, dass sich die drei Teilskalen zu einem sinnvollen (homogenen) Gesamtkonstrukt ergänzen.

Zunächst steht die Frage im Vordergrund, ob sich die für die Kodierung der transkribierten Interviewtexte gewählten Kategorien als geeignete Instrumente zur Überprüfung der faktoriellen Struktur des Konstrukts Analysekompetenz erwiesen haben. Erste Indizien für die Eignung der Kategorien wurden bereits geliefert: (1) Die vorgenommenen Berechnungen der Inter- und Intracoderreliabilität sowie der Validitätsprüfung nach Fröh (2015) erbrachten gute Werte, sodass von einer hohen Stabilität, Wiederholbarkeit und Genauigkeit der Messung ausgegangen werden kann (s. S. 136). (2) Es wurde gezeigt, dass für *jede* Kategorie hinreichend viele Kodierereignisse registriert werden konnten und sich *insgesamt* – über alle Kategorien hinweg – eine entsprechend hohe Gesamtsumme an Kodierereignissen ergab (s. Tab. 38, S. 137).

Unter Berücksichtigung der oben eingeführten Sprachregelung (die verwendeten Kategorien stellen Variablen dar, die zusammen eine homogene Teil-Skala [Faktor] des Konstrukts Analysekompetenz repräsentieren) ergibt sich nun das weitere Vorgehen: Zunächst wird die Reliabilität für die einzelnen Variablen berechnet, die die drei Teil-Skalen bilden (didaktisch-methodisches Wissen, fachliches Wissen, Verstehenstiefe). Danach werden die Korrelationen der Summen der jeweiligen Kodierereignisse angeführt und vor dem Hintergrund der Summen der jeweiligen Reliabilitäten für die Teilskalen interpretiert. In einem weiteren Schritt wird geprüft, ob die Daten normalverteilt vorliegen. Sollten diese nicht dem Kriterium der Normalverteilung genügen, wird als *non-parametrisches* Verfahren der Mann-Whitney-Test (U-Test) gewählt, um die Gruppenunterschiede zwischen ExpertInnen und NovizInnen auf Signifikanz zu prüfen und abschließend die erzielten Effektstärken für diese Unterschiede zu betrachten.

#### 5.4.1 Reliabilitätsanalysen der bisherigen Ergebnisse

In diesem Teilabschnitt stelle ich zunächst die für die Teilskalen ermittelten Reliabilitätswerte dar, um zu klären, ob mit den zur Kodierung verwendeten Variablen überhaupt eine hinreichende Messgenauigkeit erzielt werden konnte. Hinreichende Reliabilitätswerte stellen eine zentrale Voraussetzung für die daran anschließende Beurteilung der Validität dar, also für die Beantwortung der Frage, ob sich die erhaltenen Daten im Sinne der faktoriellen Struktur des Konstrukts Analysekompetenz interpretieren lassen.<sup>86</sup> Tabelle 46 gibt zunächst einen Überblick über die Anzahl der Variablen und über die errechneten Reliabilitätswerte für jede der drei Teilskalen.

<sup>86</sup> Bei der Interpretation von Trennschärfen werden Werte zwischen .4 und .7 als „angemessen“ bzw. „gut“ aufgefasst (s. z.B. Döring & Bortz, 2016, S. 478; Kelava & Moosbrugger, 2012, S. 87; Pospeschill, 2010, S. 82; Fisseni, 1997).

**Tab. 46:** Überblick über die erhaltenen (vorläufigen) Reliabilitätswerte

Faktoren des Konstrukts		Cronbach-Alpha	Anzahl der Variablen
Analysekompetenz			
Inhaltliche Dimension	Didaktisch-methodisches Wissen (DMW)	.715	12
	Fachliches (incl. fachdidaktisches) Wissen (FW)	.634	7
	<b>Gesamt (DMW + FW)</b>	<b>.818</b>	<b>19</b>
Formale Dimension	Verstehenstiefe (VT)	.619	5

Die Skala für die inhaltliche Dimension umfasst 19 Variablen, wovon 12 für das didaktisch-methodische Wissen und 7 für das fachliche Wissen stehen. Die Skala der formalen Dimension besteht aus fünf Variablen. Der für die inhaltliche Dimension erzielte Reliabilitätswert liegt bei insgesamt .818, der Wert für die formale Dimension bei .619. Innerhalb der inhaltlichen Dimension beträgt die Reliabilität für die Teilskala didaktisch-methodisches Wissen .715 und für die Teilskala fachliches (incl. fachdidaktisches) Wissen .634. Um diese Reliabilitätswerte vor dem Hintergrund des Gesamtkonstrukts Analysekompetenz bewerten zu können, ist zudem ein differenzierter Blick auf die Trennschärfewerte der einzelnen Variablen und auf ihren Beitrag zur Reliabilität der jeweiligen Skala notwendig. In Tabelle 47 sind die spezifischen Werte für die 19 Variablen der *inhaltlichen* Dimension aufgelistet.

Die Trennschärfen für die einzelnen Variablen liegen zum großen Teil zwischen .25 und .77. Als Ausnahmen fallen insbesondere drei Variablen mit geringen Trennschärfewerten auf (die Variable *Diskursivität* [Disk] mit .187, die Variable *Anwendung/Transfer* [AT] mit .163 und die Variable *Durcharbeiten von Fachbegriffen* [DFB] mit .110), sodass sich die Frage stellt, ob diese Variablen ggf. aus der Skala herausgenommen werden sollten. Zur Beantwortung dieser Frage erscheint zunächst ein Blick auf die Reliabilitäten der drei Teilskalen (didaktisch-methodisches Wissen, fachliches Wissen, Verstehenstiefe) und auf die entsprechenden Trennschärfen für die einzelnen Variablen innerhalb dieser Teilskalen hilfreich.

**Tab. 47:** (vorläufige) Reliabilitäts- und Trennschärfewerte für die 19 Variablen der inhaltlichen Dimension (didaktisch-methodisches und fachliches Wissen)

Faktoren des Konstrukts Analysekompetenz für die inhaltliche Dimension	Variablen	Korrigierte Variablen-Skala-Korrelation	Cronbach-Alpha, wenn Variable gelöscht
Didaktisch-methodisches Wissen (DMW)	Zielorientierung (ZO)	.771	.782
	Strukturierung (S)	.420	.809
	Klarheit (K)	.643	.800
	Exemplarizität (EXE)	.639	.795
	Lebensweltbezug des Themas (LWB)	.297	.815
	Verknüpfung mit dem Vorwissen (VW)	.340	.813
	Aktivierung/Motivierung zur geistigen Selbsttätigkeit (AMgS)	.254	.825
	Diskursivität des Unterrichtsprozesses (Disk)	.187	.819
	Individualisierung/Differenzierung (InDi)	.247	.817
	Ergebnissicherung/Erfolgskontrolle (EsEk)	.250	.817
	Anwendung/Transfer (AT)	.163	.820
	Lernatmosphäre (LA)	.375	.812
Fachliches (incl. fachdidaktisches) Wissen (FW)	Physikalische Begriffsbildung (PB)	.626	.795
	Durcharbeiten von Fachbegriffen (DFB)	.110	.821
	Experimentieren als zentrale Fachmethode (EZF)	.685	.793
	(weitere) fachliche Arbeitsweisen (FAW)	.482	.808
	Historische Genese physikalischen Wissens (HGPW)	.454	.812
	Produktiver Umgang mit Fehlern (PUF)	.517	.812
	Lernen als konstruktive Tätigkeit von SchülerInnen (LKS)	.267	.818

### Reliabilität der Teilskala „didaktisch-methodisches Wissen“ (DMW)

Die Reliabilität dieser Skala beträgt .715 (s. Tab. 46, S. 159). In Tabelle 48 sind die Reliabilitäts- und Trennschärfewerte für die Teilskala didaktisch-methodisches Wissen dargestellt, die aus 12 Variablen besteht.

Die beiden Variablen *Diskursivität* (Disk) und *Anwendung/Transfer* (AT) zeigen, wie sich bereits bei der Betrachtung der Skala für alle Variablen der inhaltlichen Dimension (s. Tab. 47, S. 160) andeutete, zu geringe Trennschärfewerte mit jeweils .076. Diese unbefriedigenden Werte könnten inhaltlich mit der ambivalenten Ausprägung beider Variablen im zu analysierenden Unterricht zusammenhängen. Als Folge dieser Uneindeutigkeiten ist es für die interviewten Personen schwierig, eine klare Beurteilung abzugeben. Im Einzelnen bedeutet das:

Hinsichtlich der Variable *Diskursivität* (Disk) ist im Einzelfall schwierig zu entscheiden, ob die SchülerInnen an bestimmten Stellen des Unterrichts Gelegenheit haben, das inhaltliche und methodische Vorgehen mitzubestimmen. Denn der von den ProbandInnen zu analysierende Physikunterricht ist durchgehend stark lehrerzentriert. Die wenigen Gelegenheiten, mit denen das Unterrichtsgespräch geöffnet und den SchülerInnen Möglichkeit gegeben wird, den Gang des Unterrichts mitzubestimmen, werden allzu schnell wieder durch die Lehrperson geschlossen. Deshalb stellt sich durchgehend die nicht einfach zu beantwortende Frage, ob es sich bei den entsprechenden Situationen eher um eine Scheinbeteiligung der SchülerInnen handelt und ihre Beiträge von der Lehrperson nur als Stichwortgeber für das weitere Unterrichtsgeschehen mediatisiert werden. Als Folge könnten sich bei den interviewten Personen dann Unsicherheiten in der Beurteilung der jeweiligen Situation ergeben haben.

Entsprechende Schwierigkeiten zeigen sich auch bei der Kodierung der Variable *Anwendung/Transfer* (AT). In der Schlussphase des Unterrichts präsentiert die Lehrperson eine Fresnel-Linse. Ihre Funktion ist letztlich zwar eine Anwendung des Brechungsgesetzes, das das Thema der Stunde ist, aber in den vorangegangenen Phasen des Unterrichts wurden nicht die notwendigen Voraussetzungen dafür geschaffen, dass die SchülerInnen die technische Gestaltung solcher Linsen wirklich verstehen können. Insofern kann man unterschiedlicher Meinung darüber sein, ob es sich hier wirklich um eine sinnvolle Anwendung bzw. um einen stringenten Transfer des Gelernten handelt. Auch diese Ambivalenz könnte sich in einer gewissen Unsicherheit im Urteil der interviewten Personen niederschlagen haben.

Diese inhaltlichen Gründe sprechen zusammen mit den geringen Trennschärfen dafür, die beiden Variablen aus der Teilskala didaktisch-methodisches Wissen und somit aus der Gesamtskala für die inhaltliche Dimension zu entfernen.

**Tab. 48:** (vorläufige) Reliabilitäts- und Trennschärfewerte für die Teilskala didaktisch-methodisches Wissen (inhaltliche Dimension)

<b>Faktor des Konstrukts</b> <b>Analysekompetenz für</b> <b>die inhaltliche</b> <b>Dimension</b>	<b>Variablen</b>	<b>Korrigierte</b> <b>Variablen-</b> <b>Skala-Korre-</b> <b>lation</b>	<b>Cronbach-</b> <b>Alpha, wenn</b> <b>Variable</b> <b>gelöscht</b>
Didaktisch-metho- disches Wissen (DMW)	Zielorientierung (ZO)	.751	.613
	Strukturierung (S)	.361	.695
	Klarheit (K)	.596	.670
	Exemplarizität (EXE)	.588	.658
	Lebensweltbezug des Themas (LWB)	.280	.705
	Verknüpfung mit dem Vorwissen (VW)	.463	.688
	Aktivierung/Motivierung zur geistigen Selbsttätigkeit (AMgS)	.267	.723
	Diskursivität des Unterrichts- prozesses (Disk)	.076	.726
	Individualisierung/Differenzierung (InDi)	.224	.712
	Ergebnissicherung/Erfolgskontrolle (EsEk)	.180	.717
	Anwendung/Transfer (AT)	.076	.723
	Lernatmosphäre (LA)	.358	.697



### Reliabilität der Teilskala „fachliches Wissen“ (FW)

Die Reliabilität dieser Skala beträgt .634 (s. Tab. 46, S. 159). Tabelle 49 zeigt die Reliabilitäts- und Trennschärfewerte für die Teilskala fachlichen Wissen, die aus 7 Variablen besteht.

**Tab. 49:** (vorläufige) Reliabilitäts- und Trennschärfewerte für die Teilskala fachliches Wissen (FW)

Faktor des Konstrukts Analysekompetenz für die inhaltliche Dimension	Variablen	Korrigierte Variablen-Skala- Korrelation	Cronbach- Alpha, wenn Variable gelöscht
Fachliches (incl. fach- didaktisches) Wissen (FW)	Physikalische Begriffsbildung (PB)	.549	.517
	Durcharbeiten von Fachbegriffen (DFB)	.383	.605
	Experimentieren als zentrale Fachme- thode (EZF)	.457	.560
	(weitere) fachliche Arbeitsweisen (FAW)	.454	.581
	Historische Genese physikalischen Wissens (HGPW)	.282	.626
	Produktiver Umgang mit Fehlern (PUF)	.490	.606
	Lernen als konstruktive Tätigkeit von SchülerInnen (LKS)	.164	.655

Die Trennschärfewerte für die meisten Variablen sind akzeptabel. Das gilt auch für den Wert für die Variable *Durcharbeiten von Fachbegriffen* (DFB), der innerhalb dieser Teilskala bei .383 liegt. Im Vergleich zu dessen Stellenwert in der Gesamtskala für die inhaltliche Dimension (s. Tab. 47, S. 160; Trennschärfe = .110) zeigt sich hier also, dass diese Variable hinreichend genaue Messungen zulässt. Aufgrund dessen sollte diese Variable auch in der Skala verbleiben.

Dagegen fällt der Wert für die Variable *Lernen als konstruktive Tätigkeit von SchülerInnen* (LKS) mit .164 sehr niedrig aus. Es war beabsichtigt, diese Variable für die Analyse und Kodierung der Interviews zu nutzen, weil der Sachverhalt, der durch sie definiert wird, nicht nur in der allgemeindidaktischen Diskussion eine große Rolle spielt, sondern auch in der Fachdidaktik Physik viel Beachtung gefunden hat (s. z.B. Duit, 1993, 2006; Widodo & Duit, 2004, 2005; Wiesner, 2006). Deutet man die Variable *Lernen als konstruktive Tätigkeit von SchülerInnen* lerntheoretisch, dann wäre sie eher als eine fächerübergreifende Größe anzusehen, die systematisch gesehen nicht in eine Teilskala zum fachlichen Wissen passt. Die geringe Trennschärfe deutet darauf hin, dass dies offensichtlich der Fall ist. Deshalb wird diese Variable aus der Skala fachliches Wissen und somit aus der Gesamtskala für die inhaltliche Dimension eliminiert.

### Reliabilität der Teilskala „Verstehenstiefe“ (VT)

Tabelle 50 gibt die Reliabilitäts- und Trennschärfewerte für die Teilskala Verstehens-tiefe wieder, die aus fünf Variablen besteht.

**Tab. 50:** (vorläufige) Reliabilitäts- und Trennschärfewerte für die Teilskala Verstehenstiefe (VT)

Faktor des Konstrukts Analysekompetenz für die formale Dimension	Variablen	Korrigierte Variablen-Skala-Korrelation	Cronbach-Alpha, wenn Variable gelöscht
Verstehenstiefe (VT)	Verstehenstiefe – Level 1 (VT-1)	.524	.593
	Verstehenstiefe – Level 2a (VT-2a)	.760	.313
	Verstehenstiefe – Level 2b (VT-2b)	.412	.564
	Verstehenstiefe – Level 3a (VT-3a)	.207	.646
	Verstehenstiefe – Level 3b (VT-3b)	.387	.608

Die Reliabilität der Teilskala beträgt .619 (s. Tab. 46, S. 159) und die Trennschärfewerte liegen zwischen .207 und .760. Im Falle der Variable *Verstehenstiefe – Level 3a* (VT-3a) könnte der geringe Wert (.207) mit der insgesamt geringen Anzahl an Kodierungen für diese Variable zusammenhängen (ExpertInnen = 3/NovizInnen = 0; s. Tab. 41, S. 140). Diese wenigen Kodierereignisse könnten wiederum aus dem methodischen Vorgehen bei der Präsentation des Unterrichtsvideos resultieren. Das Video wurde in vier Sequenzen unterteilt, sodass die ProbandInnen möglicherweise stärker auf die Betrachtung und Beurteilung dieser Abschnitte fokussiert waren und nur selten auf den gesamten Unterrichtsverlauf, auf den die Variable *VT-3a* zielt, Bezug nahmen.

### 5.4.2 Endgültige Ergebnisse der Qualitativen Inhaltsanalyse

Aufgrund der im vorangegangenen Abschnitt dargestellten und inhaltlich diskutierten Reliabilitäts- und Trennschärfewerte wurden aus der Teilskala didaktisch-methodisches Wissen (DMW) die beiden Variablen *Diskursivität* (Disk) und *Anwendung/ Transfer* (AT) und aus der Teilskala fachliches Wissen (FW) die Variable *Lernen als konstruktive Tätigkeit von SchülerInnen* (LKS) ausgeschlossen.

Der Ausschluss dieser Variablen hat zur Folge, dass die Anzahl der Kodierereignisse, die in die weiteren Berechnungen eingehen, entsprechend gesunken ist. Tabelle 51 stellt nun die *endgültigen* Ergebnisse der Qualitativen Inhaltsanalyse im Überblick dar.

**Tab. 51:** Endgültige Anzahl der vorgenommenen Kodierungen für die inhaltliche und formale Dimension

Faktoren des Konstrukts		ExpertInnen	Novizinnen	Gesamt
<b>Analysekompetenz</b>				
Inhaltliche Dimension	Didaktisch-methodisches Wissen (DMW)	180	97	<b>277</b>
	Fachliches (incl. fachdidaktisches) Wissen (FW)	97	44	<b>141</b>
Formale Dimension	Verstehenstiefe (VT)	134	36	<b>170</b>
<b>Summe</b>		<b>411</b>	<b>177</b>	<b>588</b>

Insgesamt liegen jetzt noch 588 Kodierereignisse vor (ursprünglich 708). Das ist immer noch eine beträchtliche Zahl, die zeigt, dass die verwendeten Variablen ein sinnvolles Instrumentarium zur Kodierung der Interviews darstellen. Auch die oben bereits erwähnten Unterschiede zwischen ExpertInnen und NovizInnen fallen immer noch deutlich aus.

Durch den Ausschluss der drei Variablen *Diskursivität* (Disk), *Anwendung/Transfer* (AT) und *Lernen als konstruktive Tätigkeit von SchülerInnen* (LKS) aus der inhaltlichen Dimension ergibt eine erneute Berechnung die in Tabelle 52 (s. S. 166) dargestellten Reliabilitäts- und Trennschärfewerte. Die Reliabilität der Skala für die inhaltliche Dimension liegt bei .824 (s. Tab. 56, S. 169). Die Trennschärfen decken für die meisten Variablen einen Wertebereich zwischen .2 und .7 ab (s. Tab. 52, S. 166). Im Vergleich zu allen anderen 15 Variablen fällt allerdings auch hier wieder der Wert für die Variable *Durcharbeiten von Fachbegriffen* (DFB) mit .021 sehr niedrig aus.

Berechnet man die Reliabilitäts- und Trennschärfewerte für die Variablen des didaktisch-methodischen Wissens und für die des fachlichen Wissens aber noch einmal getrennt (s. Tab. 53, S. 167 und Tab. 54, S. 168), dann ergibt sich für die Variable *Durcharbeiten von Fachbegriffen* (DFB) ein Trennschärfewert von .271. Aufgrund dieses akzeptablen Wertes (.271) und geschuldet der Tatsache, dass es sich bei der Variable um einen wichtigen Aspekt der zu analysierenden Unterrichtsstunde handelt, sollte diese Variable auf jeden Fall in der Gesamtskala verbleiben.

**Tab. 52:** Endgültige Reliabilitäts- und Trennschärfewerte für die Skala der inhaltlichen Dimension mit 16 Variablen

Faktoren des Konstrukts Analysekompetenz für die inhaltliche Dimension	Variablen	Korrigierte Variablen-Skala-Korrelation	Cronbach-Alpha, wenn Variable gelöscht
Didaktisch-metho- disches Wissen (DMW)	Zielorientierung (ZO)	.823	.781
	Strukturierung (S)	.507	.810
	Klarheit (K)	.636	.805
	Exemplarizität (EXE)	.669	.798
	Lebensweltbezug des Themas (LWB)	.347	.819
	Verknüpfung mit dem Vorwissen (VW)	.284	.822
	Aktivierung/Motivierung zur geistigen Selbsttätigkeit (AMgS)	.218	.837
	Individualisierung/ Differenzierung (InDi)	.205	.825
	Ergebnissicherung/ Erfolgskontrolle (EsEk)	.207	.826
	Lernatmosphäre (LA)	.416	.816
Fachliches (incl. fach- didaktisches) Wissen (FW)	Physikalische Begriffs- bildung (PB)	.605	.801
	Durcharbeiten von Fach- begriffen (DFB)	.021	.830
	Experimentieren als zentrale Fachmethode (EZF)	.638	.801
	(weitere) fachliche Arbeits- weisen (FAW)	.527	.811
	Historische Genese physika- lischen Wissens (HGPW)	.480	.817
	Produktiver Umgang mit Fehlern (PUF)	.560	.817

**Reliabilität der Teilskala „didaktisch-methodisches Wissen“ (DMW)****Tab. 53:** Endgültige Reliabilitäts- und Trennschärfewerte für die Teilskala didaktisch-methodisches Wissen mit 10 Variablen

Faktor des Konstrukts Analysekompetenz für die inhaltliche Dimension	Variablen	Korrigierte Variablen-Skala-Korrelation	Cronbach-Alpha, wenn Variable gelöscht
Didaktisch-methodisches Wissen (DMW)	Zielorientierung (ZO)	.769	.641
	Strukturierung (S)	.388	.721
	Klarheit (Kürzel: K)	.550	.704
	Exemplarizität (EXE)	.649	.677
	Lebensweltbezug des Themas (LWB)	.326	.729
	Verknüpfung mit dem Vorwissen (VW)	.446	.719
	Aktivierung/Motivierung zur geistigen Selbsttätigkeit (AMgS)	.237	.763
	Individualisierung/Differenzierung (InDi)	.162	.744
	Ergebnissicherung/Erfolgskontrolle (EsEk)	.145	.750
	Lernatmosphäre (LA)	.442	.713

In Tabelle 53 sind in der mittleren Spalte die korrigierten Trennschärfewerte für die einzelnen Variablen aufgeführt. Sie liegen bei acht Variablen zwischen .237 und .769. Lediglich zwei Variablen zeigen einen Wert von .145 (*Ergebnissicherung/Erfolgskontrolle* [EsEk]) und .162 (*Individualisierung/Differenzierung* [InDi]). Dennoch sollten diese beiden Variablen nicht aus der Skala eliminiert werden, weil durch sie Aspekte effektiven Lehrens zur Sprache gebracht werden, die mit den anderen Kategorien nicht erfasst werden können und die insofern die Aussagekraft der Skala inhaltlich stützen.

### Reliabilität der Teilskala „fachliches Wissen“ (FW)

**Tab. 54:** Endgültige Reliabilitäts- und Trennschärfewerte für die Teilskala fachliches Wissen mit 6 Variablen

Faktor des Konstrukts Analysekompetenz für die inhaltliche Dimension	Variablen	Korrigierte Variablen-Skala-Korrelation	Cronbach-Alpha, wenn Variable gelöscht
Fachliches (incl. fachdidaktisches) Wissen (FW)	Physikalische Begriffsbildung (PB)	.612	.508
	Durcharbeiten von Fachbegriffen (DFB)	.271	.651
	Experimentieren als zentrale Fachmethode (EZF)	.437	.592
	(weitere) fachliche Arbeitsweisen (FAW)	.497	.607
	Historische Genese physikalischen Wissens (HGPW)	.251	.657
	Produktiver Umgang mit Fehlern (PUF)	.617	.614

### Reliabilität der Teilskala „Verstehenstiefe“ (VT)

Da aus der Skala Verstehenstiefe (formale Dimension) keine Variablenstreichungen vorgenommen wurden, sind die entsprechenden Reliabilitäts- und Trennschärfewerte identisch geblieben. In Analogie zu den endgültigen Werten für die Skala der inhaltlichen Dimension (s. Tab. 52, S. 166) sind die entsprechenden Werte noch einmal in der folgenden Tabelle 55 aufgeführt.

**Tab. 55:** Endgültige Reliabilitäts- und Trennschärfewerte für die Skala der formalen Dimension mit 5 Variablen

Faktor des Konstrukts Analysekompetenz für die formale Dimension	Variablen	Korrigierte Variablen-Skala-Korrelation	Cronbach-Alpha, wenn Variable gelöscht
Verstehenstiefe (VT)	Verstehenstiefe – Level 1 (VT-1)	.524	.593
	Verstehenstiefe – Level 2a (VT-2a)	.760	.313
	Verstehenstiefe – Level 2b (VT-2b)	.412	.564
	Verstehenstiefe – Level 3a (VT-3a)	.207	.646
	Verstehenstiefe – Level 3b (VT-3b)	.387	.608

Tabelle 56 zeigt abschließend die endgültigen Reliabilitätswerte für die inhaltliche und formale Dimension (bzw. für die drei Teilskalen):

**Tab. 56:** Überblick über die endgültigen Reliabilitätswerte

Faktoren des Konstrukts		Cronbach-Alpha	Anzahl der Variablen
Analysekompetenz			
Inhaltliche Dimension	Didaktisch-methodisches Wissen (DMW)	.740	10
	Fachliches (incl. fachdidaktisches) Wissen (FW)	.655	6
	<b>Gesamt (DMW + FW)</b>	<b>.824</b>	<b>16</b>
Formale Dimension	Verstehenstiefe (VT)	.619	5

Die Ergebnisse aus Tabelle 56 sind eine Bestätigung für Hypothese H10: Die Reliabilitätsanalysen belegen, dass die 21 Variablen sowohl in inhaltlicher Hinsicht eine hinreichend genaue Messung der Aktivierung didaktisch-methodischen Wissens und fachlichen (incl. fachdidaktischen) Wissens ermöglichen als auch eine hinreichend genaue Erfassung der Verstehenstiefe in formaler Hinsicht (s. Tab. 52 bis Tab. 55, S. 166ff.).

Die Reliabilität der 16 Variablen umfassenden Skala der *inhaltlichen* Dimension beträgt .824. Die Reliabilitäten für die beiden Teilskalen der inhaltlichen Dimension (didaktisch-methodisches und fachliches Wissen) liegen bei .740 (zehn Variablen) bzw. .655 (sechs Variablen). Das sind mit Blick auf die Anzahl der zugehörigen Variablen immer noch akzeptable Werte. Die Reliabilität der fünf Variablen umfassenden Skala der *formalen* Dimension (Verstehenstiefe) beträgt .619. Dieser Wert liegt zwar deutlich unter dem der Skala für die inhaltliche Dimension, aber in Anbetracht der Tatsache, dass diese Skala mit nur fünf Variablen wesentlich kürzer ausfällt, ist der Wert noch akzeptabel. Zudem ist in Betracht zu ziehen, dass die Zahl der auswertbaren Kodierereignisse (s. Tab. 38, S. 137 bzw. Tab. 51, S. 165) bei den NovizInnen mit 36 Ereignissen sehr niedrig liegt und dass in der Kategorie *Verstehenstiefe – Level 3b* nur drei Kodierereignisse überhaupt identifiziert werden konnten (ExpertInnen = 3/NovizInnen = 0; s. Tab. 45, S. 145). Das ist eine schmale Datenbasis für die Berechnung von Reliabilitäten.

Zusätzliche Evidenz für die Annahme der Homogenität der Gesamtskala und dafür, dass sich die drei Teilskalen zu einem sinnvollen Gesamtkonstrukt ergänzen, ergibt sich aus den *Korrelationen* zwischen den Teilsummen für die jeweiligen Kodierereignisse (s. die oben in Tab. 51, S. 165) angegebenen Teilsummen für das didaktisch-methodische und fachliche Wissen und für die Verstehenstiefe). Die entsprechenden Korrelationswerte sind in der folgenden Tabelle 57 (s. S. 170) dargestellt.<sup>87</sup> Die Korrelationen der einzelnen Teilsummen untereinander fallen durchweg hoch aus.

<sup>87</sup> Es handelt sich hier um Pearson-Korrelationen, da die ermittelten Daten intervallskaliert sind und angenommen wird, dass deren Zusammenhang linear interpretiert werden kann (s. z.B. Pospeschill, 2006). Es werden also keine Korrelationen nach Spearman berichtet, die auf Rangunterschieden beruhen. Diese Entscheidung für die Pearson-Korrelationen ist unabhängig von der Prüfung der Signifikanz der Unterschiede zwischen ExpertInnen und NovizInnen mit dem Mann-Whitney-Test (U-Test), über den weiter unten berichtet wird. Der U-Test wird gewählt, weil die vorliegenden intervallskalierten Daten aufgrund der kleinen Stichprobe (N = 18) nicht die Bedingungen für einen T-Test erfüllen. Der Kolmogorow-Smirnow-Test zeigt zwar, dass

**Tab. 57:** Korrelationen der Summen der Kodierereignisse der drei Teilskalen untereinander und mit der Gesamtsumme aller Kodierereignisse

	Summe didaktisch- methodisches Wissen (DMW)	Summe fachliches (incl. fachdidaktisches) Wissen (FW)
Summe fachliches (incl. fachdidaktisches) Wissen (FW)	.682**	
Summe Verstehenstiefe (VT)	.561*	.663**

An dieser Stelle sei aber darauf verwiesen, dass solch eine Korrelation *keine* kausalen Beziehungen abbildet, sodass nicht gesagt werden kann, dass bspw. das didaktisch-methodische Wissen ein Prädiktor für die Verstehenstiefe ist. Die Korrelationen zeigen „nur“ Beziehungen zwischen den Skalen an: In diesem Fall handelt es sich um einen *positiven* Zusammenhang, d.h. eine Person, die oft ihr didaktisch-methodisches Wissen anwendet, aktiviert auch oft ihr fachliches (incl. fachdidaktisches) Wissen und erbringt eine tiefere Analyseleistung (abgebildet durch die erreichte Verstehenstiefe) bei der zu analysierenden Unterrichtsstunde. Somit bestätigen die Korrelationswerte die Konsistenz der drei Skalenwerte.

### 5.4.3 Beurteilung der Unterschiede zwischen ExpertInnen und NovizInnen

Die vorangegangene Betrachtung der Reliabilitäten der Teilskalen und der Gesamtskala konnte zeigen, dass die verwendeten Variablen eine hinreichend genaue Messung der Wissensaktivierung von ExpertInnen und NovizInnen (in inhaltlicher wie formaler Hinsicht) ermöglichen. Insofern stellt dieses Ergebnis eine wichtige Vorbedingung für die weiteren Betrachtungen der Ergebnisse der Qualitativen Inhaltsanalyse dar, wobei nun der Schwerpunkt auf der Frage liegt, ob und inwiefern sich ExpertInnen und NovizInnen in der Analyse von Unterricht und des dabei aktivierten Wissens unterscheiden.

**Tab. 58:** Mittelwerte, Standardabweichungen und Standardfehler für die ermittelten Kodierereignisse nach Gruppen und Teilskalen

	Status	N	Mittelwert	Standard- abweichung	Standardfehler Mittelwert
Summe didaktisch- methodisches Wissen (DMW)	ExpertInnen	9	20,00	8,337	2,779
	NovizInnen	9	10,67	3,000	1,000
Summe fachliches (incl. fachdidakti- sches) Wissen (FW)	ExpertInnen	9	10,78	4,206	1,402
	NovizInnen	9	4,69	4,076	1,359
Summe Verstehens- tiefe (VT)	ExpertInnen	9	14,89	2,759	,920
	NovizInnen	9	4,00	2,345	,782

für die Skalen DM und VT annähernd normalverteilte Werte vorliegen und die Varianzhomogenität gegeben ist, für die Teilskala FW ist dies aber bei den NovizInnen nicht der Fall (s. Tab. 59, S. 171).



Tabelle 51 (s. S. 165) zeigt die über die Qualitative Inhaltsanalyse für die ExpertInnen und NovizInnen ermittelten Kodierereignisse. Auf den ersten Blick sind deutliche Unterschiede zwischen ExpertInnen und NovizInnen zu sehen. In Ergänzung zu Tabelle 51 gibt Tabelle 58 die Mittelwerte und Standardabweichungen für die ermittelten Kodierereignisse der Gruppen (ExpertInnen vs. NovizInnen) und Teilskalen wieder.

Um nun zu überprüfen, ob es sich hier um *signifikante* Gruppenunterschiede handelt, ist zunächst zu klären, mit welchem Testverfahren diese Unterschiede geprüft werden sollen. Die in Tabelle 51 (s. S. 165) angeführten Kodierereignisse können, wie bereits oben erläutert, als intervallskalierte Daten betrachtet werden. Diese stammen allerdings aus einer kleinen Stichprobe ( $N = 18$ ), sodass fraglich ist, ob die Werte der untersuchten Variablen normalverteilt sind. Nur dann könnte die statistische Auswertung mit parametrischen Verfahren erfolgen. Zur Beantwortung der Frage der Normalverteilung wurde zunächst der Kolmogorow-Smirnow-Test durchgeführt. Dabei ergaben sich die in Tabelle 59 angeführten Ergebnisse.

**Tab. 59:** Kolmogorow-Smirnow-Test auf Normalverteilung

Status		Kolmogorow-Smirnow		
		Statistik	df	Sig.
Summe didaktisch-methodisches Wissen (DMW)	1 ExpertInnen	.133	9	.200*
	2 NovizInnen	.155	9	.200*
Summe fachliches (incl. fachdidaktisches) Wissen (FW)	1 ExpertInnen	.260	9	.081
	2 NovizInnen	.163	9	.200*
Summe Verstehenstiefe (VT)	1 ExpertInnen	.182	9	.200*
	2 NovizInnen	.278	9	.044

Der Kolmogorow-Smirnow-Test zeigt, dass hinsichtlich der von den ExpertInnen erzielten Summe beim Faktor fachliches Wissen und hinsichtlich der von den NovizInnen erzielten Summe beim Faktor Verstehenstiefe *keine* Normalverteilung vorliegt.

Aufgrund der nicht gegebenen Normalverteilung für diese Werte wurde anschließend der Mann-Whitney-Test (U-Test) durchgeführt (s. Bortz & Lienert, 2008; Bortz, Lienert, & Boehnke, 2008; Lienert, 1975; Rost, 2013). Tabelle 60 (s. S. 172) zeigt die Werte für die *Signifikanzprüfung*.

**Tab. 60:** Mann-Whitney-Test (U-Test)<sup>88</sup>

	Summe didaktisch- methodisches Wissen (DMW)	Summe fachliches (incl. fachdidakti- sches) Wissen	Summe Verstehenstiefe (VT)
Mann-Whitney-Test (U-Test)	12,500	12,500	.000
Wilcoxon-W	57,500	57,500	45,000
z	-2,481	-2,483	-3,599
Asymp. Sig. (2-seitig)	.013	.013	.000
Exakte Sig. [2*(1-seitige Sig.)]	.011b	.011b	.000b

Die Ergebnisse der Signifikanzprüfung (untere Zeile der Tab. 60) bestätigen Hypothese H11: ExpertInnen und NovizInnen unterscheiden sich hinsichtlich aller drei Faktoren (didaktisch-methodisches und fachliches Wissen, Verstehenstiefe) *signifikant* voneinander (0,01%-Niveau).

Diese Unterschiede schlagen sich auch in entsprechenden *Effektstärken* nieder (s. Tab. 61). Die Effektstärke ist beim Mann-Whitney-Test (U-Test) definiert durch die Formel: Z-Wert dividiert durch Wurzel aus N (s. Fritz, Morris, & Richler, 2012). Demnach ergeben sich hinsichtlich der einzelnen Teilsummen und für die Gesamtsumme folgende Effektstärken:

**Tab. 61:** Effektstärken

	Effektstärken
Didaktisch-methodisches Wissen (DMW)	.58
Fachliches (incl. fachdidaktisches) Wissen (FW)	.58
Verstehenstiefe (VT)	.85

Für die Beurteilung der über den Mann-Whitney-Test (U-Test) errechneten Effektstärken stellen nach Cohen (1992, S. 157, auch 1988) Werte  $\leq .1$  einen kleinen, Werte um .3 einen mittleren und Werte  $\geq .5$  einen hohen Effekt dar. Da sämtliche Werte größer als .5 sind, kann nach Cohen von *hohen* Effekten gesprochen werden; die ermittelten Unterschiede zwischen den ExpertInnen und NovizInnen sind also auch von *praktischer Bedeutsamkeit*.

<sup>88</sup> a. Gruppierungsvariable: Status; b. nicht für Bindungen korrigiert

## 6. Zusammenfassung

Nachdem im vorangegangenen Kapitel ausführlich über die Ergebnisse meiner Untersuchung berichtet wurde, sollen in diesem Kapitel die wesentlichen Ergebnisse, unter Berücksichtigung der methodischen Anlage, vorgestellt werden, um dadurch die zugrunde liegende Fragestellung der Arbeit abschließend zu beantworten.

Die vorliegende Arbeit ging im Wesentlichen der Frage nach, ob sich die faktorielle Struktur des Konstrukts Analysekompetenz<sup>89</sup>, wie sie von Plöger und Scholl (2014a) theoretisch postuliert und von Plöger et al. (2015) auf quantitativem Wege überprüft wurde, auch auf qualitativem Wege replizieren lässt, d.h. ob sich die manifesten Daten, die ich mit meiner qualitativen Studie erhoben habe, als Belege für die Struktur des Konstrukts Analysekompetenz interpretieren lassen. Konkreter formuliert: Lässt sich durch die Auswertung von Interviews (mit ExpertInnen und NovizInnen) nachweisen, dass eine angemessene Analyse von Unterricht nur dann gelingt, wenn das Wissen der betreffenden Personen eine hohe *inhaltliche* und *formale* Qualität besitzt? Dies wäre dann der Fall, wenn die Personen zum einen über ein hinreichendes didaktisch-methodisches und fachliches (incl. fachdidaktisches) Wissen und zum anderen über einen entsprechend hohen Grad der Komplexität der Informationsverarbeitung verfügen.

Den Ausgangspunkt meiner Untersuchung bildeten zwei Validierungsbelege, die von Plöger et al. (2015) und Plöger und Scholl (2014a) erbracht wurden:

- 1) In der Untersuchung von Plöger et al. (2015) sollte der Frage der Validität des Konstrukts Analysekompetenz auf quantitativem Wege nachgegangen werden. Die zentrale Annahme dabei lautete, dass das Konstrukt Analysekompetenz durch zwei Dimensionen, eine inhaltliche und eine formale, bestimmt ist. Die inhaltliche Dimension setzt sich aus zwei Faktoren zusammen: dem didaktisch-methodischen und dem fachlichen (incl. fachdidaktischen) Wissen. Die formale Dimension bildet den Grad der für die Analyseleistung notwendigen Komplexität der Informationsverarbeitung ab. Im Rückgriff auf das fünfstufige Kompetenzmodell ist diese formale Dimension durch drei Faktoren definiert: Analytische Kompetenz (Stufen 1 und 2 des Modells), Synthetische Kompetenz (Stufe 3 und 4) und Prozesskompetenz (Stufe 5) (s. Kap. 1.4.2). Um die durch diese Annahmen postulierte Struktur des Konstrukts zu prüfen, wurden die Dimensionen und zugehörigen Faktoren auf unterschiedliche Weise kombiniert, sodass sich dadurch fünf mögliche Modelle konstruieren ließen. Konfirmatorische Faktorenanalysen (CFA) zeigten, dass sich nur dann hinreichende Fit-Werte ergaben, wenn das betreffende Modell die postulierte zweidimensionale Struktur des Konstrukts (incl. aller Faktoren) spiegelte (s. Kap. 1.6.2). Die Ergebnisse der von Plöger, Scholl und Seifert durchgeführten CFA sprechen für die Validität des Konstrukts Analysekompe-

---

89 Dabei verstehen Plöger und Scholl unter Analysekompetenz die Fähigkeit, „die Qualität ... von beobachtetem Unterricht im Hinblick auf dessen Lernwirksamkeit angemessen erfassen und bewerten zu können“ (Plöger & Scholl, 2014a, S. 88).

tenz und stellen die Referenz für den von mir durchgeführten qualitativen Validierungsversuch dar. Mit Blick auf die erzielten Fit-Werte (s. Tab. 3, S. 46) hatte ich die Option, zwischen Modell 4 und 5 zu wählen, denn beide Modelle stellen adäquate Repräsentationen des Konstrukts dar. Meine Entscheidung ist letztlich zugunsten des fünften Modells ausgefallen, damit – in Analogie zur Studie von Plöger et al. (2015) – der *subitizing*-Faktor erfasst werden konnte.

- 2) Der zweite Validierungsbeleg von Plöger und Scholl (2014a) basierte auf dem sogenannten well-known-group-Vergleich (Cronbach & Meehl, 1955; DeVellis, 1991; Hattie & Cooksey, 1984; Rubin & Babbie, 2015, s. dazu Kap. 1.6, S. 41 & Kap. 4, S. 131). Dessen Grundgedanke zielt auf den Vergleich zweier Gruppen, von denen man vermutet, dass das in Frage stehende Merkmal (hier Analysekompetenz) bei den Personen der einen Gruppe besonders hoch und bei den Personen der anderen Gruppe besonders niedrig ausgeprägt ist. Entsprechen die Testergebnisse diesen erwarteten Unterschieden, so ist dies ein wichtiges Indiz für die Validität des zugrunde liegenden Konstrukts Analysekompetenz. Die von Plöger und Scholl erhobenen Daten zeigten, dass Fach- und SeminarleiterInnen (ExpertInnen) durchweg die höchsten Stufen der Analysekompetenz erreichten, die Studierenden (NovizInnen) dagegen nur die niedrigeren Stufen. Zwischen diesen beiden Gruppen lagen die Analyseleistungen der ReferendarInnen und praktizierenden Lehrpersonen (s. Kap. 1.5.2, S. 40 sowie Plöger und Scholl, 2014a, S. 104). Diese großen Unterschiede zwischen den ExpertInnen und den NovizInnen waren mit Blick auf die Erkenntnisse der Expertiseforschung „erwartbar“ und ließen sich somit als weiteren Beleg für die Validität des zugrunde liegenden Konstrukts Analysekompetenz interpretieren.

Im Sampling meiner Untersuchung waren daher Personen vertreten, die einerseits Unterricht sehr gut analysieren konnten, und andererseits Personen, die über weniger gute Analyseleistungen verfügten. Da sich mit Blick auf die Ausprägung der Analysekompetenz im Projekt von Plöger und Scholl gezeigt hatte, dass Studierende nur relativ geringe Analyseleistungen erbrachten, Fach- und SeminarleiterInnen dagegen sehr hohe (s. Plöger & Scholl, 2014a), wurden für meine Studie Studierende als NovizInnen und Fach- und SeminarleiterInnen als ExpertInnen rekrutiert. Insgesamt bestand das Sampling aus 18 Personen, also aus jeweils neun ExpertInnen und neun NovizInnen.<sup>90</sup>

Den ProbandInnen wurde als *Stimulus* eine videographierte Unterrichtsstunde vorgeführt (11. Klasse, Fachoberschule, Fach Physik; Thema: Snelliussches Brechungsgesetz). Dieses Video wurde ausgewählt, damit sich die Aussagen der ProbandInnen (im Fragebogen von Plöger et al., 2015 und im Interview) auf den *identischen* Stimulus, also auf dieselbe Unterrichtsstunde, bezogen und somit die Vergleichbarkeit der beiden Validierungsversuche gegeben war (s. Kap. 2.2).

---

<sup>90</sup> Darüber hinaus wurde auch der fachliche Hintergrund der ProbandInnen berücksichtigt, denn Plöger und Scholl (2014a) konnten zeigen, dass Personen, die das Fach Physik bzw. ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichteten bzw. studierten, signifikant bessere Analyseleistungen erbrachten als Personen, die keines dieser Fächer vertraten. Daher waren im Sampling meiner Untersuchung eben diese Personen vertreten: Personen, die das Fach Physik unterrichten bzw. studieren (Kürzel: Ph), Personen, die ein anderes naturwissenschaftliches Fach und/oder Mathematik unterrichten bzw. studieren (Kürzel: NW) sowie Personen, die keines dieser Fächer vertreten (Kürzel: KNW) (s. Kap. 2.3.4, S. 57f.).

Auf Grundlage dieser Unterrichtsstunde wurden *fokussierte, leitfadengestützte Interviews* geführt (s. z.B. Fuhs, 2007; Helfferich, 2011). Diese wurden mit einem Tonbandgerät aufgezeichnet und anschließend *vollständig* transkribiert (s. z.B. Langer, 2010). Um die transkribierten Interviews auszuwerten und die faktorielle Struktur des Konstrukts Analysekompetenz abbilden zu können, habe ich mich für die *Qualitative Inhaltsanalyse* entschieden (s. z.B. Mayring, 2014, 2015). Das Ziel einer Qualitativen Inhaltsanalyse besteht darin, Texte systematisch zu analysieren (s. z.B. Mayring, 2002). Diese *systematische* Analyse soll die Textaussagen ordnen, die Texte also *strukturell* verdichten, auf wesentliche Aussagen reduzieren und somit „auf den Begriff bringen“. Diese Begriffe werden in der Sprache der Qualitativen Inhaltsanalyse als *Kategorien* bezeichnet (Mayring, 2015, S. 51). Die entsprechenden Kategorien wurden nicht – wie vielfach üblich – auf induktivem Wege ermittelt, sondern *deduktiv* entwickelt, sodass sie die beiden Dimensionen (incl. ihrer Faktoren) aus dem Validierungsversuch von Plöger et al. (2015) abbildeten. Die jeweiligen Kategorien setzten sich aus einer Definition, einer Beschreibung und einem Ankerbeispiel zusammen (s. Kap. 3).

Die Kodierung der Interviews erfolgte im *Konsensverfahren* mit vier Personen (s. Kap. 2.5.3). Dabei wurden alle Interviews dreimal ausgewertet: Zunächst wurden die Interviews unter Berücksichtigung der Kategorien für das didaktisch-methodische Wissen analysiert, anschließend unter dem Gesichtspunkt des fachlichen (incl. fachdidaktischen) Wissens und abschließend unter dem Aspekt der Verstehenstiefe. Textstellen, die sich mit diesen Kategorien kodieren ließen, belegten, dass die interviewten Personen didaktisch-methodisches und fachliches (incl. fachdidaktisches) Wissen bei der Analyse aktivierten und die zu analysierende Unterrichtsstunde auf unterschiedlich komplexe Weise verarbeiteten.

Die im Konsensverfahren festgestellten Textstellen wurden selbst noch einmal in Frage gestellt, um so zu überprüfen, ob die ausgewählten Textstellen auch von externen Kodierenden identifiziert werden konnten. Dazu wurden zwei externe Kodierenden gebeten, nach einem entsprechenden Trainingsverfahren, ca. 10% der Kodierungen zu überprüfen. Die Überprüfung sollte zeigen, ob die Kodierungen reproduzierbar waren und ein hohes Maß an intersubjektiver Übereinstimmung aufwiesen.

Im Idealfall hätten die Kodierenden spezifische Textstellen auf die gleiche Weise kodieren müssen, wie diejenigen Personen, die bereits im vorangegangenen Konsensverfahren beteiligt waren. Zur Überprüfung der Text-Kategorien-Zuordnung wurden drei Verfahren gewählt (s. Kap. 2.5.4 & Kap. 2.5.5): (1) Zunächst wurde die *Intracoderreliabilität* gemessen. Ziel war es zu prüfen, ob ein und dieselbe Kodierende zu zwei Zeitpunkten eine Textstelle mit der gleichen Kodierung versieht (s. Lauf, 2001; Mayring, 2010, 2015). Die Absicherung dieser *individuellen* Stabilität in der Kodierung stellte eine entscheidende *Voraussetzung* für die darüber hinausgehende Interdecoderreliabilität dar. (2) Zur Ermittlung der *Interdecoderreliabilität* wurden die Kodierungen der beiden Kodierenden miteinander verglichen und überprüft, ob sie „nachweislich zu ähnlichen Ergebnissen“ (Mayring, 2009, S. 471) gekommen sind. Der Idealfall läge hier also vor, wenn *beide* Kodierenden ein und derselben Textstelle ein und dieselbe Kategorie zugeordnet hätten. (3) Abschließend erfolgte eine *Validitätsprüfung*, bei der die Kodierungen der externen Kodierenden mit den im Konsensverfahren markierten Textstellen verglichen wurden. Dieses Vorgehen entspricht der Validitätsprüfung nach Fröh (2015): Wenn die externen Kodierenden Textstellen mit der gleichen Kategorie versehen wie die Personen des Konsensverfahrens,

dann teilen *alle* ein gemeinsames Verständnis des Kategoriensystems und – davon kann bei Übereinstimmung hinsichtlich der Kategorisierungen ebenfalls ausgegangen werden – eine *gemeinsame* theoretische Basis, die in diesem Fall aus der zweidimensionalen Struktur des Konstrukts Analysekompetenz bestand. Die Auswertung dieser drei unterschiedlichen Verfahren ergab gute Cohens-Kappa-Werte, die alle im und über dem anvisierten Zielbereich von .6 bis .75 lagen (s. Tab. 37, S. 136). Diese Werte belegten, dass die vorgenommenen Kodierungen wiederholbar, stabil und zuverlässig waren.

Die kodierten Textstellen dokumentierten die Häufigkeiten der von Gruppen (z.B. ExpertInnen vs. NovizInnen) und Subgruppen (z.B. ExpertInnen-Physik; NovizInnen-Physik) aktivierten Wissensfacetten. Auf diese Weise ließen sich entsprechende Unterschiede deutlich machen, wie sie sich z.B. in der grundsätzlichen Überlegenheit bei der Wissensaktivierung des didaktisch-methodischen Wissens der ExpertInnen zeigten. Mithilfe der Ergebnisse ließen sich auch eine Reihe von Hypothesen prüfen, die die faktorielle Validität des Konstrukts Analysekompetenz betrafen (s. Kap. 5.2).

Insgesamt gesehen erwies sich die genutzte Methode der Qualitativen Inhaltsanalyse bei Erhebung und Interpretation der Daten als hilfreiches Instrument. Darüber hinaus zeigten sich aber auch die Grenzen dieses methodischen Zugriffs in dreifacher Hinsicht:

- 1) Es blieb ungeklärt, ob es sich bei den augenscheinlich deutlichen Unterschieden zwischen ExpertInnen und NovizInnen um statistisch bedeutsame, also signifikante Unterschiede handelte.
- 2) Aus den Darstellungen zur Häufigkeiten der Wissensaktivierung wurde nicht ersichtlich, ob *einzelne* Personen die gesamte Palette der Facetten (definiert durch die entsprechenden Kategorien) nutzten oder vorwiegend bestimmte Wissensfacetten aktivierten und andere Facetten möglicherweise seltener oder gar nicht. Wenn sich aber zeigen ließe, dass die einzelnen Personen nicht nur einige, sondern im Prinzip alle Facetten des jeweiligen Faktors der inhaltlichen Dimension (didaktisch-methodisches Wissen oder fachliches Wissen) genutzt haben, um die Unterrichtsstunde zu analysieren, dann wäre dies ein Beleg für die inhaltliche Kohärenz der Facetten *innerhalb eines Faktors*. Darin würde ich dann auch eine nachträgliche Rechtfertigung der Wahl der Kategorien für den jeweiligen Faktor sehen, deren Legitimation aus den Forschungen über Prinzipien effektiven Lehrens, aus fachdidaktischen Argumentationen und Ergebnissen der Expertiseforschung hergeleitet wurden (s. Kap. 1.2 bis Kap. 1.4 & Kap. 3).
- 3) Die Anzahl der Kodierungen für didaktisch-methodisches Wissen einerseits und für fachliches (incl. fachdidaktisches) Wissen andererseits ließ auch keine Schlüsse darüber zu, ob Personen, die häufiger (seltener) Facetten der einen Wissensart (z.B. didaktisch-methodisches Wissen) aktivierten, auch häufiger (seltener) Facetten der anderen Wissensart (z.B. fachliches Wissen) nutzten. Wäre dies der Fall gewesen, so wäre dies ein Indiz für die Kohärenz *zwischen den beiden Faktoren* der inhaltlichen Dimension und somit auch dafür, dass sich beide Wissensarten bei der Analyse von Unterricht effektiv ergänzen.

Daher wurden die ermittelten Wissensaktivierungen weiter quantifiziert, um Signifikanzprüfungen, Reliabilitätsanalysen sowie Trennschärfe- und Korrelationsberechnungen durchführen zu können. Dazu habe ich mich an den Begrifflichkeiten für Reliabilitätsanalysen orientiert (s. Kap. 4, S. 133f. & Kap. 5.4, S. 156ff.). Im Rahmen von Reliabilitäts-

analysen sollen Testitems eine hinreichende Reliabilität und eine möglichst hohe Korrelation mit anderen Items der betreffenden Skala aufweisen. Dabei repräsentiert die Reliabilität die Messgenauigkeit der Items und entsprechende Korrelationen weisen darauf hin, dass die Items eine homogene Skala bzw. Teilskala bilden. In diesem Sinne wurden die Kategorien als *Variablen* bezeichnet. Das Kategoriensystem, das in Kapitel 3 vorgestellt wurde, beinhaltete demnach in der inhaltlichen Dimension zwölf Variablen für den Faktor des didaktisch-methodischen Wissens und sieben Variablen für den Faktor des fachlichen (incl. fachdidaktischen) Wissens. In der formalen Dimension umfasste der Faktor der Verstehenstiefe fünf Variablen. Dabei wurden die drei Faktoren, in Entsprechung zu Itemskalen, als Skalen bzw. Teilskalen des Konstrukts Analysekompetenz interpretiert.

Die spezifische Ausprägung der Variablen ergab sich daraus, wie oft eine Person Wissen aktivierte, welches einer bestimmten Variable zugeordnet werden konnte. Jede dieser Wissensaktivierungen wurde dabei als ein „Treffer“ gewertet, für den, wie bei einer richtigen Antwort eines Testitems, ein Punkt vergeben wurde. Entsprechende Differenzen zwischen den ProbandInnen konnten so als entsprechende Differenzen hinsichtlich der Ausprägung der Wissensaktivierungen verstanden werden. Darüber hinaus konnten diese *intervallskalierten* Daten genutzt werden, um zum einen mit entsprechenden Reliabilitätsanalysen überprüfen zu können, ob die ermittelten Kodierungen als ausreichend genaue Messungen über alle Variablen interpretiert werden können. Zum anderen konnten die vermuteten Unterschiede in den Wissensaktivierungen zwischen ExpertInnen und NovizInnen auf statistische Signifikanz geprüft werden.

Mit Blick auf quantitative Weiterverarbeitung der Daten ergaben sich folgende zentrale Ergebnisse:

- 1) Die Ergebnisse meiner Untersuchung lieferten eindeutige Belege, mit denen die zweifaktorielle Struktur des Konstrukts Analysekompetenz, wie von Plöger und Scholl (2014a) theoretisch modelliert und von Plöger et al. (2015) mittels CFA nachgewiesen wurde, auch auf qualitativem Wege mittels der hier angewandten Qualitativen Inhaltsanalyse repliziert werden konnte (Generalhypothese 1).
- 2) Die im Sinne des *well-known-group-Vergleichs* vermuteten Unterschiede zwischen ExpertInnen und NovizInnen bei der Analyse von Unterricht konnten ebenfalls repliziert werden und lieferten somit eine Bestätigung für die Generalhypothese 2.
- 3) Insgesamt wurden 588 Textstellen kodiert, die für die Aktivierung des Wissens der ProbandInnen stehen (s. Tab. 51, S. 165). Dies ist eine achtbare Anzahl, die zeigt, dass die verwendeten Variablen ein sinnvolles Instrumentarium zur Kodierung der Interviews darstellen.
- 4) Weiterhin konnte der Nachweis über eine hinreichend genaue Messung der Aktivierung didaktisch-methodischen und fachlichen (incl. fachdidaktischen) Wissens in inhaltlicher Hinsicht sowie eine hinreichend genaue Erfassung der Verstehenstiefe in formaler Hinsicht erbracht werden. Das zeigen die durchgehend guten bzw. akzeptablen Reliabilitätswerte in Relation zur Länge der Teilskalen (s. Tab. 56, S. 169): Der Reliabilitätswert für die *inhaltliche* Dimension beträgt .824 für die 16 Variablen umfassende Skala. Die Teilskalen des didaktisch-methodischen und des fachlichen (incl. fachdidaktischen) Wissens weisen mit .740 und .655 akzeptable Reliabilitätswerte auf. Hinsichtlich der *formalen* Dimension ergab sich ein Reliabilitätswert von .619 für die fünf Variablen umfassende Skala. Dabei gilt zu berücksichtigen, dass die Kategorisierung von Textstellen, in denen die ProbandInnen ihr Wissen aktivierten, hier durch einen



Akt der Interpretation erfolgte. Im Vergleich zu einer methodisch standardisierten Vorgehensweise (wie z.B. bei der quantitativen Messung der Analysekompetenz im Test von Plöger und Scholl, 2014a) liegt hier also ein gewisser *interpretativer* Spielraum vor, der sich mindernd auf die Reliabilität der Messungen ausgewirkt haben könnte.

- 5) Die Ergebnisse des Mann-Whitney-Tests (U-Test) zeigten signifikante Gruppenunterschiede zwischen den ExpertInnen und NovizInnen (s. Tab. 60, S. 172). Die Werte haben gezeigt, dass sich ExpertInnen und NovizInnen hinsichtlich aller drei Faktoren signifikant unterscheiden.
- 6) Diese Unterschiede schlugen sich auch in entsprechenden Effektstärken nieder (für das didaktisch-methodisches Wissen: .58; für das fachliche [incl. fachdidaktische] Wissen: .58; für die Level der Verstehenstiefe: .85, s. dazu Tab. 61, S. 172). Da sämtliche Werte größer als .5 waren, kann nach Cohen (1988) von *hohen* Effekten gesprochen werden. Die ermittelten Unterschiede zwischen den ExpertInnen und NovizInnen sind somit auch von praktischer Bedeutsamkeit.



## 7. Ausblick

Zum Abschluss meiner Arbeit möchte ich einen kurzen Ausblick darauf geben, welche Perspektiven sich daraus für weitere Forschungen ergeben könnten. Dabei beschränke ich mich exemplarisch auf zwei Aspekte, nämlich (1) auf die Möglichkeiten des Mixed-Methods-Ansatz zur Prüfung der Validität von Konstrukten (s. Kap. 7.1) und (2) auf die bisher weitgehend ungeklärte Frage zwischen didaktisch-methodischem und fachlichem Wissen, wie sie insbesondere von Shulman im Sinne seiner „Amalgamthese“ formuliert wurde (s. Kap. 7.2).

### 7.1 Das Potenzial des Mixed-Methods-Ansatzes zur Konstruktvalidierung

Das Ziel der vorliegenden Arbeit war es, die zweifaktorielle Struktur des Konstrukts Analysekompetenz, wie es von Plöger und Scholl (2014a) theoretisch modelliert und von Plöger et al. (2015) durch Konfirmatorische Faktorenanalysen (CFA) nachgewiesen wurde, auf qualitativem Wege mittels Qualitativer Inhaltsanalyse zu replizieren. Beide Forschungsansätze waren auf die Frage gerichtet, ob sich die auf unterschiedlichen Wegen generierten Daten – dort quantitativ, hier qualitativ – mit den theoretischen Annahmen des zugrunde liegenden Konstrukts Analysekompetenz kohärent interpretieren lassen. Im Ergebnis zeigte sich: Sowohl die quantitativen als auch die qualitativen Daten lassen sich stringent mit den Annahmen interpretieren, durch die das Konstrukt Analysekompetenz definiert ist.

Diese Ergänzung von quantitativen und qualitativen Verfahren ist bisher eher selten zur Validierung von Konstrukten eingesetzt worden (s. Daigneault & Jacob, 2014). Künftig sollte man der Kombination dieser unterschiedlichen Ansätze im Sinne des Mixed-Methods-Ansatzes weitaus mehr Raum geben. Das Potenzial dieses Ansatzes sehe ich nicht nur darin, zu übereinstimmenden Ergebnissen hinsichtlich der Validierung zu kommen (so in meiner Arbeit), sondern auch in den Stärken und Schwächen von quantitativen und qualitativen Verfahren, durch die jeweils Facetten von Validität diskutiert und geprüft werden können, die mit dem komplementären Verfahren nur bedingt erfassbar sind, denn die Beurteilung der Validität ist auf *verschiedene Quellen* empirischer Evidenz angewiesen ist (s. AERA et al., 2014).

Wie man sich eine solche Ergänzung mit Blick auf die jeweiligen Stärken konkret vorstellen kann, sei hier in Kürze vor dem Hintergrund meiner Arbeit und dessen Bezug zum Projekt von Plöger, Scholl und Seifert (2015) erläutert. Diese Stärken werden mit Blick auf das in der Einleitung (s. S. 14f.) skizzierte Kompetenzkonzept von Blömeke et al. (2015) erkennbar, in dem Kompetenz als ein Kontinuum modelliert wird, das zwischen den Polen von *latenten Dispositionen* (kognitive und affektiv-motivationale Einstellungen) und *manifesten Leistungen* (beobachtbares Verhalten) gespannt ist:

- 1) Die Stärke des quantitativen Ansatzes erwies sich darin, dass er detaillierte Informationen über den Pol der *performances* (s. Einleitung, S. 15) lieferte. Mit dem entwickelten Testinstrument konnten in relativ kurzer Zeit eine hohe Zahl an ProbandInnen untersucht und präzise numerische Daten über die Ausprägung ihrer Analysekompetenz gesammelt werden.<sup>91</sup> Aufgrund der relativ großen Stichprobe ( $N = 800$ ) ließen sich – in Verbindung mit dem Nachweis der spezifischen Objektivität der Itemskala – vorsichtige Schlüsse hinsichtlich der Generalisierbarkeit der Testergebnisse ziehen, so dass davon ausgegangen werden kann, dass das eingesetzte quantitative Instrument grundsätzlich geeignet ist, die Analyseleistung von Lehrpersonen (zumindest die Leistung von jenen, die am Gymnasium unterrichten) zuverlässig zu diagnostizieren. Insgesamt gesehen lieferte also der Einsatz des quantitativen Messinstruments präzise Informationen über die vorliegenden Leistungen der ProbandInnen. Aus diesen Leistungen ließen sich auf indirekte Weise auch Aussagen über den anderen Pol des Kompetenzkonzepts (*dispositions*) ableiten, denn die durchgeführten Konfirmatorischen Faktorenanalysen (CFA) belegten, dass die Varianz in den Testscores sinnvoll mit der postulierten Struktur der *underlying dispositions* erklärt werden konnte.

Damit stößt der quantitative Ansatz allerdings zugleich an seine Grenzen, da er zwar die Wissensbestände der ProbandInnen erfasst, die diese bei der Lösung der Items nutzen, es aber ungeklärt bleibt, ob das Wissen, das von den ForscherInnen bei der Konstruktion der Items als notwendig zu deren Lösung unterstellt wurde, weitgehend identisch mit dem von den ProbandInnen im Test eigenständig aktivierten Wissen ist.

- 2) Bei der Beantwortung dieser Frage zeigte allerdings der qualitative Ansatz seine Stärken. Denn die geführten Interviews und deren Auswertung mittels Qualitativer Inhaltsanalyse machten die konkreten Denkprozesse der ProbandInnen sichtbar (s. Kap. 5.3). Ihre Aussagen entsprachen jenen Wissensbeständen, die durch die zwei Dimensionen und die zugehörigen Faktoren des Konstrukts Analysekompetenz definiert und über entsprechende Kategorien operationalisiert war (s. Kap. 3). Hinsichtlich der *inhaltlichen* Dimension konnte man das z.B. daran sehen, dass bei der Analyse des Unterrichts *didaktisch-methodisches* Wissen, wie z.B. Zielorientierung und Strukturierung, zur Beurteilung der Lernwirksamkeit der zu analysierenden Unterrichtsstunde aktiviert wurde. Dies gilt ebenso für die Aktivierung *fachlichen* (incl. *fachdidaktischen*) Wissens. Und hinsichtlich der *formalen* Dimension repräsentierten die Äußerungen der ProbandInnen unterschiedliche Grade der Komplexität der Informationsverarbeitung, die im Sinne des Konstrukts als verschiedene Level der Verstehenstiefe definiert waren.

Die Dokumentation solcher Wissensaktivierungen und komplexer Verstehensprozesse ist eine Stärke des qualitativen Vorgehens; mit den Testitems lassen sich dagegen nur einzelne Momente dieser Prozesse punktuell einfangen. Insgesamt gesehen konnte der qualitative Ansatz also differenzierte Informationen über den Pol der *underlying dispositions* liefern. Diese auf qualitativem Wege erzielten Ergebnisse stellen somit – zusammen mit der oben herausgestellten Stärke des quantitativen Vorgehens – eine sinnvolle Kombination von Verfahren zur Validierung von Konstrukten dar und sollten künftig intensiver in entsprechenden Forschungsprojekten genutzt werden.

---

<sup>91</sup> Bei der Analyse dieser Daten konnte die Raschhomogenität der Itemskala und damit auch die spezifische Objektivität der Itemparameter nachgewiesen werden. Dadurch war sichergestellt, dass der (relative) Schwierigkeitsgrad der Items über alle vier untersuchten Teilpopulationen hinweg weitgehend identisch war.

Dieses Mixed-Methods Vorgehen muss allerdings durch weitere Bemühungen zum Nachweis von Validität ergänzt werden. Unerlässlich sind darüber hinaus z.B. *experimentelle Designs*, durch die auf systematische Weise Informationen gewonnen werden können, damit die Wirksamkeit spezifischer Kompetenzfacetten – definiert durch die jeweiligen Konstrukte – evaluiert werden kann. Ein solches Vorgehen folgt der Forschungslogik: „Wenn sich theoretisch begründen lässt, dass bestimmte experimentell variierende Faktoren einen *Effekt auf ein Konstrukt* haben sollten, dann sollte eine experimentelle Variation dieser Faktoren einen Effekt auf die Testwerte haben, die das Konstrukt repräsentieren.“ (Hartig et al., 2012, S. 156). Im Sinne dieser Logik ergeben sich zur Überprüfung derartiger Effekte zwei Möglichkeiten der Konstruktvalidierung:

- 1) Zum einen könnte ein Treatment zur Verbesserung des didaktisch-methodischen Wissens konzipiert werden. Solch eine Interventionsstudie in Form eines *quasi-experimentellen Designs*, könnte Auskunft darüber geben, ob und in welchem Maße sich die Aneignung des didaktisch-methodischen Wissens auf die Ausprägung der Analysekompetenz auswirkt. Plöger und Scholl (2016) konnten bereits zeigen, dass ein entsprechendes Treatment positiven Einfluss auf „die Entwicklung der didaktisch-methodischen Kompetenzen und damit der Analysekompetenz“ (Plöger & Scholl, 2016, S. 192) der ProbandInnen hatte.
- 2) Zum anderen könnte, im Anschluss an diese bereits vorliegende Interventionsstudie, ein Treatment zur Verbesserung des fachlichen Wissens entwickelt werden, mit dem überprüft werden kann, ob auch eine Verbesserung des fachlichen Wissens positive Auswirkungen auf die Analysekompetenz hat. Ein solches Treatment wird derzeit durchgeführt und kommt in Kürze zum Abschluss.

Diese Interventionsstudien stellen zusammen mit den beiden hier berichteten Studien ein *erweitertes* Forschungsprogramm dar. Mit solch einem Programm begibt man sich zwar auf einen langen und steinigen Weg, aber dadurch können durch Nutzung unterschiedlicher Perspektiven und Methoden plausible, empirisch gesicherte Argumente für ein Konstrukt gesammelt werden (Blömeke, 2007; Cronbach, 1988; Kane, 2013).

## 7.2 Ein Ausblick auf die „Amalgamthese“ Shulmans

Die im Rahmen meiner Arbeit erzielten Ergebnisse besitzen auch eine Relevanz für die Forschung zum *pedagogical content knowledge* (PCK, fachdidaktisches Wissen). Die dortige Diskussion um die Natur des PCK dauert nun bereits rund drei Jahrzehnte an. Die Anfänge dieser Diskussion gehen auf die Arbeiten von Shulman zurück (1986, 1987) und wurzelten in seiner Sorge, dass die damaligen Bestrebungen in der Ausbildung von Lehramtsstudierenden vorrangig auf die Vermittlung von allgemeinen pädagogischen Fähigkeiten (general pedagogical skills) gerichtet waren und dass gleichzeitig Fragen, die den Inhalt einer Unterrichtsstunde und somit das Professionswissen von Lehrpersonen betrafen, zu sehr vernachlässigt wurden. Vor diesem Hintergrund forderte er, dass in der künftigen Ausbildung von Lehrpersonen wieder stärker die Inhalte betont sowie Fachwissen und pädagogisches Wissen als „part of one indistinguishable body“ (Shulman, 1986, S. 6) gesehen werden sollten.

Die Begründung dieser Forderung war verbunden mit dem Entwurf einer Klassifikation des Lehrerwissens, in der Shulman sieben Wissenskomponenten unterschied: *general pedagogical knowledge* (allgemeines, fächerübergreifendes pädagogische Wissen), *subject matter content knowledge* (Fachwissen), *pedagogical content knowledge* (fachdidaktisches Wissen), *curricular knowledge* (curriculares Wissen), *knowledge of learners and their characteristics* (Vorwissen der SchülerInnen), *knowledge of educational contexts* (Wissen über den Bildungskontext) sowie *knowledge of educational ends, purposes, and values, and their philosophical and historical grounds* (Wissen über pädagogische Ziele, Zwecke) (s. Shulman, 1987, S. 8).

Im Rahmen der verschiedenen Wissenskomponenten billigte Shulman dem PCK allerdings eine besondere Rolle zu:

Among those categories, pedagogical content knowledge is of special interest because it identifies the distinctive bodies of knowledge for teaching. It represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented and adapted to the diverse interests and abilities of learners, and presented for instruction. (Shulman, 1987, S. 8)

Im Zusammenhang mit dieser Definition stand Shulmans „Amalgamthese“: PCK ist „that specific amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding“ (ebd.). Diese Amalgamthese und die Klassifikation des Wissens nach Shulman stellten den Ausgangspunkt der Debatte über die Natur des PCK, die bis zum heutigen Tag andauert (s. z.B. Abell, 2007, 2008; Friedrichsen, van Driel, & Abell, 2011; Gess-Newsome, 2015). In dieser Diskussion lassen sich zwei zentrale Forschungsstränge unterscheiden: 1) Die Erweiterung des PCK-Konzepts von Shulman „by modifying the constituent components or adding new components“ (Park & Chen, 2012, S. 924) und (2) Untersuchungen zu den Zusammenhängen zwischen den einzelnen Komponenten.

Mit Blick auf diese beiden Perspektiven würde ich meine Arbeit der zweiten Richtung zuordnen, weil sie Hinweise darauf gibt, wie man sich künftige Forschung zum Zusammenhang zwischen entsprechenden Wissenskomponenten vorstellen kann. Denn in inhaltlicher Hinsicht ging es bei meiner Fragestellung darum, ob die Aussagen der ProbandInnen sowohl deren didaktisch-methodisches wie deren fachliches Wissen repräsentierten. Damit ist im Grunde genommen dann auch die Frage impliziert, ob diese Aussagen als eine Verschmelzung von didaktisch-methodischem Wissen (PK) und fachlichem Wissen (CK) und somit im Sinne der Amalgamthese Shulmans gedeutet werden könnten.

Das scheint mir der Fall zu sein, denn wenn man die Interviewaussagen unter dieser Perspektive betrachtet, zeigt sich, dass die einzelnen Elemente von PK und CK während der Analyse des Unterrichts nicht isoliert als „reines“ generisch-pädagogisches Wissen oder als „reines“ Fachwissen in Erscheinung treten, sondern dass sie im *situationsspezifischen* Kontext der Unterrichtsstunde *synchron* aktiviert werden: Bei der Verbalisierung der Facetten von PK treten die Facetten des CK, bei der Aktivierung des CK dagegen das PK in den Hintergrund. Diese Synchronisation könnte als Indikator für die Existenz von PCK und somit auch als Beleg für Shulmans Amalgamthese verstanden werden.

Beispielhaft sei dies an zwei ExpertInnen-Aussagen verdeutlicht. In Beispiel 1 steht das PK im Vordergrund und das CK im Hintergrund, während in Beispiel 2 das CK in den Vordergrund und das PK in den Hintergrund tritt:

### Beispiel 1: Lebensweltbezug des Themas

E: „Und jetzt müsste eigentlich von [der] Schülerseite die Frage kommen: Mein Onkel, der hat so eine dicke Brille. Warum kann der im Gestell nicht so eine Folie verwenden? Also, ... Physik ist ja nun wirklich ein Fach, was erlebbare Umwelt, erlebbare Natur, beschreibt. Dass muss man auch nutzen.“ (s. Kap. 5.3.1.1, S. 148)

Die interviewte Person merkt an, dass das Wissen, das die SchülerInnen erwerben sollen, nicht auf weitere (andere) alltägliche Situationen angewendet wird (PK), denn die Gesetzmäßigkeit wird nicht ausreichend dekontextualisiert (CK). Das bedeutet, dass die Lehrperson es versäumt, *sinnstiftendes Lernen* seitens der SchülerInnen zu initiieren, da sie keinen Lebensweltbezug herstellt (PK) und sie letztlich dadurch auch die spätere Anwendung der betreffenden Inhalte auf alltagsnahe Situationen erschwert (CK).

### Beispiel 2: Physikalische Begriffsbildung

E: [Die Schüler] „können keine Lösungsstrategien entwickeln. Das hängt auch mit seiner begrifflichen Unexaktheit zusammen ... von der ich glaube, dass er selbst nicht über ein so profundes Fachwissen verfügt, denn jemand, der wirklich das internalisiert hat, der geht mit den Begrifflichkeiten viel klarer und sauberer um und fordert die auch in der Klarheit von den Schülern ein.“ (s. Kap. 5.3.1.2, S. 150)

In diesem Beispiel unterstellt die interviewte Person der Lehrperson, dass diese nicht über profundes Fachwissen (CK) verfügt und sie dadurch nicht in der Lage ist, Klarheit (PK) bei der Verwendung der physikalischen Fachtermini von den SchülerInnen einzufordern. Diese Aussage bezieht sich auf die Forderung der KMK (2004a), dass eine angemessene Kommunikation über Sachverhalte unter physikalischen Aspekten nur dann gelingt, wenn die entsprechenden Fachbegriffe klar und verständlich eingeführt wurden (CK). Diese Forderung steht wiederum in Verbindung mit der Auffassung, dass Lernen ein Prozess der Informationsverarbeitung ist und die notwendigen Informationen von den SchülerInnen nur verstanden werden können, wenn diese mit hinreichender Klarheit (PK) formuliert sind.

Anhand dieser Beispiele zeigt sich zudem, was momentan als *Minimalkonsens* in der Diskussion um PCK feststellbar ist (s. z.B. Banks, Leach, & Moon, 2005; Drechsler & van Driel, 2008; Friedrichsen et al., 2011; Gess-Newsome, 2015; Kind, 1999; Lederman, Gess-Newsome, & Latz, 1994; Nilsson & Loughran, 2012; Park & Oliver, 2008): PK und CK sind zwei Wissenskomponenten, die das Denken und Handeln von Lehrpersonen bestimmen, wenn sie über konkreten Unterricht reflektieren.

Über diesen Minimalkonsens hinaus verweisen diese beiden Beispiele auch auf die Auffassung vieler ForscherInnen, dass die wesentliche Funktion des PCK (hier als Zusammenspiel von PK und CK verstanden) darin liegt, SchülerInnen bedeutungsvolles Lernen (*meaningful learning*) zu ermöglichen (s. z.B. Kind, 2009; Nilsson & Loughran, 2012; Shulman 1987). Sowohl in diesen beiden Beispielen als auch in den vielen anderen registrierten Aktivierungen von PK und CK wird erkennbar, dass die ExpertInnen mit Hilfe ihres Wissens das im Video beobachtbare Verhalten der Lehrperson daraufhin beurteilen, ob es zu entsprechenden *Lernerfolgen* auf Seiten der SchülerInnen geführt hat.



# Literaturverzeichnis

- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Ledermann (Eds.), *Handbook of research on science education* (pp. 1105–1149). Mahwah: Lawrence Erlbaum.
- Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, 30(10), 1405–1416.
- Aebli, H. (1971). *Grundformen des Lehrens. Ein Beitrag zur psychologischen Grundlegung der Unterrichtsmethode* (7. Aufl.). Stuttgart: Klett.
- Aebli, H. (1993a). *Zwölf Grundformen des Lehrens. Eine Allgemeine Didaktik auf psychologischer Grundlage. Medien und Inhalte didaktischer Kommunikation, der Lernzyklus* (7. Aufl.). Stuttgart: Klett.
- Aebli, H. (1993b). *Denken, das Ordnen des Tuns. Band 1: Kognitive Aspekte der Handlungstheorie* (2. Aufl.). Stuttgart: Klett-Cotta.
- AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education) (Eds.). (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education) (Eds.). (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Atteslander, P. (2010). *Methoden der empirischen Sozialforschung* (13. Aufl.). Berlin: Schmidt.
- Ausubel, D. P. (1968). *Educational psychology: a cognitive view*. New York: Holt, Rinehart and Winston.
- Ausubel, D. P., Novak, J. D., & Hanesian, H. (1980). *Psychologie des Unterrichts, Band 1* (2. Aufl.). Weinheim: Beltz.
- Baird, J. R., & Mitchel, I. J. (Eds.). (1986). *Improving the quality of teaching and learning: An Australian case study – the PEEL Project*. Melbourne: The Monash University Press.
- Baird, J. R., & Northfield, J. R. (1992). *Learning from the Peel experience*. Melbourne: Monash University Printing Services.
- Balderjahn, I. (2003). Validität. Konzepte und Methoden. *WiSt – Wirtschaftswissenschaftliches Studium*, 32(3), 130–135.
- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.
- Banks, F., Leach, J., & Moon, B. (2005). Extract from ‘New understandings of teachers’ pedagogic knowledge’. *Curriculum Journal*, 16(3), 331–340.
- Barnhart, T., & van Es, E. A. (2015). Learning to analyze teaching: developing pre-service science teachers’ abilities to notice, analyze and respond to student thinking. *Teaching and Teacher Education*, 45, 83–93.
- Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, R. Bromme & R. Lehmann (Hrsg.), *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Band 2: Mathematische und physikalische Kompetenzen in der Oberstufe* (S. 271–315). Opladen: Leske + Budrich.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J., & Kunter, M. (2011a). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Baumert, J., & Kunter, M. (2011b). Das mathematikspezifische Wissen von Lehrkräften, kognitive Aktivierung im Unterricht und Lernfortschritte von Schülerinnen und Schülern. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 163–192). Münster: Waxmann.



- Baumert, J., & Kunter, M. (2013). The COACTIV Model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Kraus, & M. Neubrand (Eds.), *Cognitive Activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV Project* (pp. 25–48). New York: Springer.
- Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.
- Bayrhuber, H., Bögeholz, S., Eggert, S., Elster, D., Grube, C., Höble, C., Linsner, M., Lücken, M., Mayer, J., Mittelsten Scheid, N., Müller, A., Nerdel, C., Neuhaus, B., Precht, H., Sandmann, A., Schmiemann, P., & Schoormans, G. (2007). Biologie im Kontext. Erste Forschungsergebnisse. *Der Mathematische und Naturwissenschaftliche Unterricht*, 60, 304–313.
- Bell, T. (2007). Entdeckendes und forschendes Lernen. In S. Mikelskis-Seifert & T. Rabe (Hrsg.), *Physik Methodik. Handbuch für die Sekundarstufe I und II* (S. 70–81). Berlin: Cornelsen Scriptor.
- Bennet, J., Hogarth, S., & Lubben, F. (2003). *A systematic review of the effects of context-based and Science-Technology-Society (STS) approaches in the teaching of secondary science. Version 1.1. Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Berliner, D. C. (1987). Der Experte im Lehrerberuf: Forschungsstrategien und Ergebnisse. *Unterrichtswissenschaft*, 15(3), 295–305.
- Berliner, D. C. (1992). The nature of expertise in teaching. In A. Oser & J.-L. Patry (Eds.), *Effective and responsible teaching: The new synthesis* (pp. 227–248). San Francisco: Jossey Bass.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463–482.
- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science Technology Society*, 24, 200–212.
- Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 56(3), 205–213.
- Bernholt, S. (2010). *Kompetenzmodellierung in der Chemie. Theoretische und empirische Reflexion am Beispiel des Modells hierarchischer Komplexität*. Berlin: Logos.
- Blomberg, G., Stürmer, K., & Seidel, T. (2011). How pre-service teachers observe teaching on video: Effects of viewers' teaching subjects and the subject of the video. *Teaching and Teacher Education*, 27(7), 1131–1140. doi: 10.1016/j.tate.2011.04.008
- Blomberg, G., Renkl, A., Sherin, G., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for Educational Research Online*, 5(1), 90–114.
- Blömeke, S. (2007). Qualitativ – quantitativ, induktiv – deduktiv, Prozess – Produkt, national – international. Zur Notwendigkeit multikriterialer und multiperspektivischer Zugänge in der Lehrerbildungsforschung. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programm-evaluation* (S. 13–36). Münster: Waxmann.
- Blömeke, S. (2013). *Validierung als Aufgabe im Forschungsprogramm „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“* (KoKoHs Working Papers, 2). Berlin: Humboldt-Universität & Johannes Gutenberg-Universität.
- Blömeke, S., & König, J. (2010). Messung des pädagogischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser, & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 239–263). Münster: Waxmann.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3–13.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerbildung*. Münster: Waxmann.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.). (2010a). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich*. Münster: Waxmann.



- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.). (2010b). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S., Suhl, U., Kaiser, G. (2011). Teacher Education Effectiveness: Quality and Equity of Future Primary Teachers' Mathematics and Mathematics Pedagogical Content Knowledge. *Journal of Teacher Education*, 62(2), 154–171.
- Blömeke, S., Kaiser, G., Döhrmann, M., & Lehmann, R. (2010). Mathematisches und mathematikdidaktisches Wissen angehender Sekundarstufen-I-Lehrkräfte im internationalen Vergleich. In S. Blömeke, G. Kaiser, & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 197–238). Münster: Waxmann.
- Blömeke, S., Kaiser, G., Lehmann, R., König, J., Döhrmann, M., Buchholtz, C., & Hacke, S. (2009). TEDS-M: Messung von Lehrerkompetenzen im internationalen Vergleich. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus, & R. Mulder (Hrsg.), *Lehrerprofessionalität – Bedingungen, Genese, Wirkungen und Messung* (S. 181–210). Weinheim: Beltz.
- Blömeke, S., Seeber, S., Lehmann, R., Kaiser, G., Schwarz, B., Felbrich, A., & Müller, C. (2008). Messung des fachbezogenen Wissens angehender Mathematiklehrkräfte. In S. Blömeke, G. Kaiser, & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 49–88). Münster: Waxmann.
- Blum, W., Krauss, S., & Neubrand, M. (2011). COACTIV – Ein mathematikdidaktisches Projekt? In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 329–343). Münster: Waxmann.
- Bogert, N. van den, Bruggen, J. van, Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teacher and Teacher Education*, 37, 208–216.
- Bogner, A., & Menz, W. (2009). Das theoriegenerierende Experteninterview. Erkenntnisinteresse, Wissensformen, Interaktion. In A. Bogner, B. Littig, & W. Menz (Hrsg.), *Experteninterviews. Theorien, Methoden, Anwendungsfelder* (3. Aufl.) (S. 61–98). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bogner, A., Littig, B., & Menz, W. (2014). *Interviews mit Experten. Eine praxisorientierte Einführung*. Wiesbaden: Springer.
- Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: a multidimensional perspective. *Learning and Instruction*, 13(3), 327–347.
- Bönsch, M. (2011). *Heterogenität und Differenzierung. Gemeinsames und differenziertes Lernen in heterogenen Lerngruppen*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Borko, H. (2004). Professional development and teacher learning: mapping the terrain. *Educational Researcher*, 33(8), 3–15.
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473–498.
- Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, 24(2), 417–436. doi:10.1016/j.tate.2006.11.012
- Born, M. (1969). *Experiment und Theorie in der Physik*. Mosbach: Physik Verlag.
- Borowski, A., & Riese, J. (2010). Physikalisch-fachdidaktisches Wissen – Was kommt in der Praxis an? *Praxis der Naturwissenschaften – Physik in der Schule*, 59(5), 5–8.
- Borowski, A., Kirschner, S., Liedtke, S., & Fischer, H. E. (2011). Vergleich des Fachwissens von Studierenden, Referendaren und Lehrenden in der Physik. *PhyDid A – Physik und Didaktik in Schule und Hochschule*, 10(1), 1–9.
- Borowski, A., Fischer, H. E., Olszewski, J., Reinhold, P., Riese, J., & Vogelsang, C. (2010). Professionswissen von Physiklehrkräften. In D. Höttercke (Hrsg.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomenen und Systematik*. Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik in Dresden 2009 (S. 365–367). Münster: LIT-Verlag.
- Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D., ... Sumfleth, E. (2010). Professionswissen von Lehrkräften in den Naturwissenschaften (ProwiN) – Kurzdarstellung des BMBF-Projekts. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 341–349.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.

- Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (3. Aufl., unter Mitarbeit von T. Barskova, K. Leitner, & R. Oesterreich). Heidelberg: Springer.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
- Bortz, J., Lienert, G. A., & Boehne, K. (2008). *Verteilungsfreie Methoden in der Biostatistik* (3. Aufl.). Heidelberg: Springer.
- Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 470–497). New York: Cambridge University Press.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687–699.
- Bromme, R. (1985). Was sind Routinen im Lehrerhandeln? Eine Begriffsklärung auf der Grundlage neuerer Ergebnisse der Problemlöseforschung. *Unterrichtswissenschaft*, 2, 182–192.
- Bromme, R. (1992). *Der Lehrer als Experte*. Göttingen: Huber.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule* (S. 172–212). Göttingen: Hogrefe.
- Bromme, R. (2001). Teacher Expertise. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 15459–15465). Oxford: Pergamon.
- Bromme, R. (2008). Lehrerexpertise. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 159–167). Göttingen: Hogrefe.
- Bromme, R., & Haag, L. (2008). Forschung zur Lehrerpersönlichkeit. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (2. Aufl., S. 803–819). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bromme, R., & Rheinberg, F. (2006). Lehrende in Schulen. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (5. Aufl., S. 296–334). Weinheim: Beltz.
- Brophy, J. (1991). *Teachers' Knowledge of Subject Matter as it relates to their Teaching Practice*. Greenwich, Cn: JAI Press.
- Brophy, J. (2000). *Teaching*. Educational practices series 1. Brüssel: International Academy of Education. Abgerufen von: <http://www.ibe.unesco.org>
- Brophy, J. (Ed.). (2004). *Advances in research on teaching: Vol 10. Using video in teacher education*. Oxford, UK: Elsevier.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Bryman, A. (2008). Of methods and methodology. *Qualitative research in organizations and management: An international Journal*, 3(2), 159–168.
- Buck, P., & Mackensen, M. (2006). *Naturphänomene erlebend verstehen*. Köln: Aulis.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. C. (1988). Expert-Novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education*, 39, 25–31.
- Carter, K., Sabers, D., Cushing, K., Pinnegar, S., & Berliner, D. C. (1987). Processing and using information about students: A study of expert, novice, and postulant teachers. *Teaching and Teacher Education*, 3(2), 147–157.
- Castro, A., Clark, K., Jacobs, J., & Givvin, K. B. (2005). Response to theory & practice question: using video to support teacher learning. *AMTE Connections*, 14(3), 8–12.
- Chi, M. T. H. (2006). Two approaches to the study of expert's characteristics. In K. A. Ericsson, N. Charnes, P. Verltovich & R. Hoffmann (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21–30). New York: Cambridge University Press.
- Chi, M. T. H., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist*, 47(3), 177–188.
- Chinn, C. A., & Brewer, W. F. (1998). Theories of knowledge acquisition. In B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education, Part One* (pp. 97–113). Dordrecht: Kluwer Academic Publishers.
- Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen. Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122–141.
- Clausen, M., Schnabel, K., & Schröder, S. (2002). Konstrukte der Unterrichtsqualität im Expertenurteil. *Unterrichtswissenschaft*, 30(3), 246–277.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Cohors-Fresenborg, E., & Kaune, C. (2003). Unterrichtsqualität: Die Rolle von Diskursivität für "guten" gymnasialen Mathematikunterricht. In H.-W. Henn (Hrsg.), *Beiträge zum Mathematikunterricht 2003*. Vorträge auf der 37. Tagung für Didaktik der Mathematik vom 3. bis 7. März 2013 in Dortmund (S. 173–180). Hildesheim: Franzbecker.
- Cohors-Fresenborg, E., & Kaune, C. (2007). Kategorisierung von Diskursen im Mathematikunterricht bezüglich metakognitiver und diskursiver Anteile. In A. Peterkop & A. Bikner-Ahsbals (Hrsg.), *mathematische Bildung – mathematische Leistung* (S. 233–248). Hildesheim: Franzbecker.
- Collins, A. M., Greeno, J. G., & Resnick, L. B. (2001). Educational learning theory. In N. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 4276–4279). Oxford: Elsevier.
- Commons, M. L. (2007). Introduction of the model of hierarchical complexity. *Behavioral Development Bulletin*, 13, 1–6.
- Commons, M. L., & Richards, F. A. (1984a). A general model of stage theory. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations, Vol. 1: Late adolescent and adult cognitive development* (pp. 120–140). New York: Praeger.
- Commons, M. L., & Richards, F. A. (1984b). Applying the general stage model. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations, Vol. 1: Late adolescent and adult cognitive development* (pp. 141–157). New York: Praeger.
- Commons, M. L., & White, M. S. (2003). A complete theory of tests for a theory of mind must consider hierarchical complexity and stage. *Behavioral and Brain Science*, 26(5), 606–607.
- Commons, M. L., Miller, P. M., & Kuhn, D. (1982). The Relation between formal operational reasoning and academic course selection and performance among college freshmen and sophomores. *Journal of Applied Developmental Psychology*, 3(1), 1–10.
- Commons, M. L., Goodheart, E. A., Dawson, T. L., & Draney, K. (2008). Using Rasch Scaled Stage Scores to Validate Orders of Hierarchical Complexity of Balance Beam task Sequences. *Journal of Applied Measurement*, 9(2), 182–199.
- Commons, M. L., Trudeau, E. J., Stein, S. A., Richards, F. A., & Krause, S. R. (1998). Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review*, 18(3), 237–278.
- Creswell, W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2. Aufl.). Thousand Oaks, CA: Sage.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Daehler, K. R., & Shinohara, M. (2001). A complete circuit is a complete circle: Exploring the potential of case materials and methods to develop teachers' content knowledge and pedagogical content knowledge of science. *Research in Science Education*, 31(2), 267–288.
- Daigneault, P.-M., & Jacob, S. (2014). Unexpected but most welcome: Mixed methods for the validation and revision of the participatory evaluation measurement instrument. *Journal of Mixed Methods Research*, 8(1), 6–24.
- de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and Qualities of Knowledge. *Educational Psychologist*, 31(2), 105–113.
- Dellinger, A. B., & Leech, N. L. (2007). Toward a unified validation framework in mixed methods research. *Journal of Mixed Methods Research*, 1(4), 309–332.
- DeVellis, R. F. (1991). *Scale development: Theory and application*. Newbury Park, CA: Sage.
- Diekmann, A. (2009). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen* (20. Aufl.). Reinbek: Rowohlt.
- Dijk, E. M. van, & Kattmann, U. (2010). Evolution im Unterricht: Eine Studie über fachdidaktisches Wissen von Lehrerinnen und Lehrern. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 7–21.
- Dilthey, W. (1979). *Gesammelte Schriften. VII. Band: Der Aufbau der Welt in den Geisteswissenschaften* (7. Aufl.). Stuttgart: Vandenhoeck & Ruprecht.
- Dittmar, N. (2004). *Transkription* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ditton H. (2002). Unterrichtsqualität – Konzeptionen, methodische Überlegungen und Perspektiven. *Unterrichtswissenschaft*, 30(3), 197–212.

- Döhrmann, M., Kaiser, G., & Blömeke, S. (2010). Messung des mathematischen und mathematikdidaktischen Wissens: Theoretischer Rahmen und Teststruktur. In S. Blömeke, G. Kaiser, & R. Lehmann, R. (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehegender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 169–196). Münster: Waxmann.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl., unter Mitarbeit von S. Pöschl). Berlin: Springer.
- DPG (Deutsche Physikalische Gesellschaft) (Hrsg.). (2006). *Thesen für ein modernes Lehramtsstudium im Fach Physik*. Bad Honnef: Deutsche Physikalische Gesellschaft.
- Drechsler, M., & van Driel, J. (2008). Experienced teachers' pedagogical content know-ledge of teaching acid-base chemistry. *Research in Science Education*, 38(5), 611–631.
- Dresing, T., & Pehl, T. (2010). Transkription. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 723–733). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Driel, J. van, & Berry, A. (2012). Teacher professional development focusing on pedagogical content knowledge. *Educational Researcher*, 41(1), 26–28.
- Driel, J. van, & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21(11), 1141–1153.
- Driel, J. van, Verloop, N., & Vos, W. de. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, 35(6), 673–695.
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit. Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Münster: Waxmann.
- Duit, R. (1989). Vorstellung und Experiment. Von der eingeschränkten Überzeugungskraft experimenteller Beobachtung. *Naturwissenschaften im Unterricht – Physik/Chemie*, 37, 37–39.
- Duit, R. (1993). Schülervorstellungen – von Lerndefiziten zu neuen Unterrichtsansätzen. *Naturwissenschaften im Unterricht – Physik*, 4(16), 4–10.
- Duit, R. (1994). An Schülervorstellungen anknüpfend Physik lehren und lernen. *Naturwissenschaften im Unterricht – Physik*, 5(22), 4–6.
- Duit, R. (1995a). Zur Rolle der konstruktivistischen Sichtweise in der naturwissenschaftsdidaktischen Lehr- und Lernforschung. *Zeitschrift für Pädagogik*, 41(6), 905–923.
- Duit, R. (1995b). Empirische physikdidaktische Unterrichtsforschung. *Unterrichtswissenschaft*, 25(2), 98–106.
- Duit, R. (1996). Lernen als Konzeptwechsel im naturwissenschaftlichen Unterricht. In R. Duit & C. v. Rhöneck (Hrsg.), *Lernen in den Naturwissenschaften. Beiträge zu einem Workshop an der Pädagogischen Hochschule Ludwigsburg* (S. 145–162). Kiel: IPN.
- Duit, R. (2003). Naturwissenschaftliches Arbeiten. *Unterricht Physik*, 14, 4–8.
- Duit, R. (2006). Schülervorstellungen und Lernen von Physik – Forschungsergebnisse und die Realität der Unterrichtspraxis. In R. Girwidz, M. Gläser-Zikuda, M. Laukenmann, & T. Rubitzko (Hrsg.), *Lernen im Physikunterricht* (S. 13–22). Hamburg: Kovač.
- Duit, R. (2015). Alltagsvorstellungen und Physik lernen. In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (3. Aufl., S. 657–680). Berlin: Springer.
- Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671–688.
- Duit, R., Gropengießer, H., & Stäudel, L. (Hrsg.). (2004). *Naturwissenschaftliches Arbeiten: Unterricht und Material*. Velber: Friedrich.
- Duit, R., Treagust, D. F., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Voosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629–646). New York: Routledge.
- Dulisch, F. (1985). *Lernen als Form menschlichen Handelns. Eine handlungstheoretisch analysierte Analyse selbstgesteuerter Lernprozesse aus wirtschafts- und berufspädagogischer Sicht*. Bergisch Gladbach: Hobein.
- Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, 22(3), 206–214.
- Dweck, C. S. (1996). Implicit theories as organizers of goals and behaviour. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 69–90). New York: The Guilford Press.
- Ecarius, J., & Miethe, I. (Hrsg.). (2011). *Methodentriangulation in der qualitativen Bildungsforschung*. Opladen: Verlag Barbara Budrich.
- Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Göttingen: Hogrefe.
- Einsiedler, W. (2002). Das Konzept „Unterrichtsqualität“. *Unterrichtswissenschaft*, 30(3), 194–196.
- Einsiedler, W., & Hardy, I. (2010). Kognitive Strukturierung im Unterricht: Einführung und Begriffsklärung. *Unterrichtswissenschaft*, 38(3), 194–209.

- Elio, R., & Scharf, P. B. (1990). Modelling novice-to-expert shifts in problem solving strategy and knowledge organization. *Cognitive Science*, 14(4), 579–640.
- Es, E. A. van, & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571–596.
- Es, E. A. van, & Sherin, M. G. (2008). Mathematics teachers' „learning to notice“ in the context of a video club. *Teaching and Teacher Education*, 24(2), 244–276.
- Fakis, A., Hilliam, R., Stonely, H., & Townend, M. (2014). Quantitative analysis of qualitative information from interviews: A systematic literature review. *Journal of Mixed Methods Research*, 8(2), 139–161.
- Fey, A., Gräsel, C., Puhl, T., & Parchmann, I. (2004). Implementation einer kontextorientierten Unterrichtskonzeption für den Chemieunterricht. *Unterrichtswissenschaft*, 32(3), 238–256.
- Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional knowledge of science teachers. In B. J. Fraser, K. G. Tobin, & C. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 435–448). New York: Springer.
- Fischer, H. E., Borowski, A., Kauertz, A., & Neumann, K. (2010). Fachdidaktische Unterrichtsforschung – Unterrichtsmodelle und die Analyse von Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 59–75.
- Fischer, H. E., Klemm, K., Leutner, D., Sumfleth, E., Tiemann, R., & Wirth, J. (2003). Naturwissenschaftsdidaktische Lehr-Lernforschung. Defizite und Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 179–209.
- Fischler, H. (1997). Was versteht man in der Physik unter „Teilchen“? *Naturwissenschaften im Unterricht – Physik*, 8, 9–11.
- Fischler, H. (2008). Physikdidaktisches Wissen und Handlungskompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 27–49.
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik. Mit Hinweisen zur Intervention* (2. Aufl.). Göttingen: Hogrefe.
- Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. New York: Cambridge University Press.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613–619.
- Flick, U. (2007). Zur Qualität qualitativer Forschung – Diskurse und Ansätze. In U. Kuckartz, H. Grunenberg, & T. Dresing (Hrsg.), *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis* (2. Aufl., S. 188–209). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Flick, U. (2009). *Sozialforschung. Methoden und Anwendung. Ein Überblick für die BA-Studiengänge*. Reinbek: Rowohlt.
- Flick, U. (2011a). *Qualitative Sozialforschung. Eine Einführung* (7. Aufl.). Reinbek: Rowohlt.
- Flick, U. (2011b). *Triangulation. Eine Einführung* (3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fluck, H.-R. (1996). *Fachsprachen: Einführung und Bibliographie* (5. Aufl.). Tübingen: Francke.
- Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, B. J. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11(2), 145–252.
- Frey, A. (2004). Die Kompetenzstruktur von Studierenden des Lehrberufs. Eine internationale Studie. *Zeitschrift für Pädagogik*, 50(6), 903–925.
- Frey, A. (2015). *Zum Verständnis von Validität in der quantitativ arbeitenden empirischen Bildungsforschung*. Vortrag auf der 80. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF), Göttingen.
- Friebertshäuser, B., & Langer, A. (2010). Interviewformen und Interviewpraxis. In B. Friebertshäuser, A. Langer, & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl., S. 437–455). Weinheim: Juventa.
- Friedrichs, J. (1990). *Methoden empirischer Sozialforschung* (14. Aufl.). Opladen: Westdeutscher Verlag.
- Friedrichsen, P., Driel, J. van., & Abell, S. K. (2011). Taking a closer look at science teaching orientations. *Science Education*, 95(2), 358–376.
- Fries, E., & Rosenberger, R. (1973). *Forschender Unterricht. Ein Beitrag zur Didaktik und Methodik des mathematischen und naturwissenschaftlichen Unterrichts in allgemeinbildenden Schulen, mit besonderer Berücksichtigung der Sekundarstufen*. Frankfurt am Main: Diesterweg.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology*, 141, 2–18.
- Früh, W. (2015). *Inhaltsanalyse. Theorie und Praxis* (8. Aufl.). Konstanz: UVK Verlagsgesellschaft.
- Fuhs, B. (2007). *Qualitative Methoden in der Erziehungswissenschaft*. Darmstadt: WBG (Wiss. Buchges.).
- Furr, R., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Los Angeles: Sage Publications.



- Gage, N. L. (1989). The paradigm wars and their aftermath: A “historical” sketch of research and teaching since 1989. *Educational Researcher*, 18, 4–10.
- Gagel, W. (2000). *Einführung in die Didaktik des politischen Unterrichts. Ein Studienbuch* (2. Aufl.). Opladen: Leske + Budrich.
- Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, 16, 41–67.
- Gess-Newsome, J. (1999). Pedagogical Knowledge. An Introduction and Orientation. In J. Gess-Newsome & N. G. Ledermann (Eds.), *Examining Pedagogical Content Knowledge: The construct and its implications for Science Education* (pp. 3–17). Dordrecht: Kluwer Academic Publishers.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, Friedrichsen, P., & Loughran, J. (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). New York: Routledge.
- GFD (Gesellschaft für Fachdidaktik) (Hrsg.). (2004). *Kerncurriculum Fachdidaktik. Orientierungsrahmen für alle Fachdidaktiken*. Abgerufen von: [http://www.fachdidaktik.org/cms/download.php?cat=Ver%C3%B6ffentlichungen&file=Publikationen\\_zur\\_Lehrerbildung-Anlage\\_3.pdf](http://www.fachdidaktik.org/cms/download.php?cat=Ver%C3%B6ffentlichungen&file=Publikationen_zur_Lehrerbildung-Anlage_3.pdf)
- Gilbert, J. K. (2004). Models and modelling: Routes to more authentic science education. *Journal of Science and Mathematics Education*, 2, 115–130.
- Gilbert, J. K., & Boulter, C. J. (1998). Learning science through models and modelling. In B. J. Fraser, & K. G. Tobin (Eds.), *International Handbook of Science Education, Part One* (pp. 53–66). Dordrecht: Kluwer Academic Publishers.
- Girwitz, R. (2015). Medien im Physikunterricht. In E. Kircher, R. Girwitz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (3. Aufl., S. 193–245). Berlin: Springer.
- Girwitz, R., & Berger, R. (2008). Physik im Alltag entdecken und verstehen. Wie Alltagsphysik im Unterricht physikalisches Interesse und Verstehen unterstützen kann. *Naturwissenschaften im Unterricht Physik*, 19(105/106), 4–7.
- Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen* (4. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gläser-Zikuda, M. (2015). Qualitative Auswertungsverfahren. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (2. Aufl., S. 119–130). Wiesbaden: Springer.
- Good, T., & Brophy, J. (2000). *Looking in classrooms* (8<sup>th</sup> Ed.). New York: Longman.
- Gramzow, Y. (2015). *Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion*. Berlin: Logos.
- Gramzow, Y., Riese, J., & Reinhold, P. (2013). Modellierung fachdidaktischen Wissens angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 7–30.
- Gräsel, C., Nentwig, P., & Parchmann, I. (2005). Chemie im Kontext: Curriculum development and evaluation strategies. In J. Bennett, J. Holman, R. Miller, & D. Waddington (Eds.), *Making a difference: Evaluation as tool for improving science education* (pp. 53–66). Münster: Waxmann.
- Groslight, I., Unger, C., & Jay, E. (1991). Understanding models and their use in science: Conceptions of Middle and High School students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Grossmann, P. I., & Schoenfeld, A. (2005). Teaching subject matter. In L. Darling-Hammond & J. D. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 201–231). San Francisco, CA: Jossey-Bass.
- Gruber, H. (1994). *Expertise. Modelle und empirische Untersuchungen*. Opladen: West-deutscher Verlag.
- Gruber, H. (2001). Acquisition of expertise. In N. J. Smelser, & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 5145–5150). Amsterdam: Elsevir.
- Gruber, H. (2004). Kompetenzen von Lehrerinnen und Lehrern – Ein Blick aus der Expertiseforschung. In A. Hartinger & M. Fölling-Albers (Hrsg.), *Lehrerkompetenzen für den Sachunterricht* (S. 21–33). Bad Heilbrunn: Klinkhardt.
- Gruber, H. (2008). Lernen und Wissenserwerb. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 95–104). Göttingen: Hogrefe.
- Gruber, H. & Gallenberger, W. (2006). Das Verhältnis von Unterrichtskompetenz und Berufserfolg. In F. Oser & M. Kern (Hrsg.), *Qualität der beruflichen Bildung – Eine Forschungsbaustelle* (S. 19–57). Bern: h.e.p.-Verlag.
- Gruber, H., & Renkl, A. (2000). Die Kluft zwischen Wissen und Handeln: Das Problem des trägen Wissens. In G. H. Neuweg (Hrsg.), *Wissen – Können – Reflexion: Ausgewählte Verhältnisbestimmungen* (S. 155–174). Innsbruck: StudienVerlag.

- Gruber, H., & Stöger, H. (2011). Experten-Novizen-Paradigma. In E. Kiel & K. Zierer (Hrsg.), *Basiswissen Unterrichtsgestaltung. Band 2: Unterrichtsgestaltung als Gegenstand der Wissenschaft* (S. 247–264). Baltmannsweiler: Schneider Verlag Hohengehren.
- Gruber, H., Harteis, C., & Rehl, M. (2006). Professional learning: Erfahrung als Grundlage von Handlungskompetenz. *Bildung und Erziehung*, 59(2), 193–203.
- Gruber, H., Mandl, H., & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze* (S. 139–156). Göttingen: Hogrefe.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen*. Münster: Waxmann.
- Grunenberg, H., & Kuckartz, U. (2010). Deskriptive Statistik in der qualitativen Sozialforschung. In B. Frieberthäuser, A. Langer, & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl., S. 487–500). Weinheim: Juventa.
- Hammann, M., Phan, T. T. H., Ehmer, M., & Bayrhuber, H. (2006). Fehlerfrei Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 59(5), 292–299.
- Handal, B. (2003). Teacher's mathematical beliefs: A review. *The Mathematics Education*, 13, 47–57.
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, 94(3), 638–645.
- Harlen, W. (1999). *Effective teaching of science – A review of research*. Edinburgh: SCORE.
- Harrison, A., & Treagust, D. (1996). Secondary student's mental models of atoms and molecules: Implication for teaching chemistry. *Science Education*, 80(5), 509–534.
- Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 135–163). Heidelberg: Springer.
- Hattie, J. (2003). *Teachers make a difference. What is the research evidence?* Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne. Abgerufen von: [http://research.acer.edu.au/research\\_conference\\_2003/4/](http://research.acer.edu.au/research_conference_2003/4/)
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon: Routledge.
- Hattie, J. (2013). *Lernen sichtbar machen* (überarbeitete deutschsprachige Ausgabe von Visible Learning, übersetzt und bearbeitet von W. Beywl und K. Zierer). Baltmannsweiler: Schneider Verlag Hohengehren.
- Hattie, J., & Cooksey, R. W. (1984). Procedures for Assessing the Validities of Tests Using the “Known-Groups” Method. *Applied Psychological Measurement*, 8(3), 295–305.
- Hattie, J., & Wollenschläger, M. (2014). A conceptualization of feedback. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen: theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 135–149). Münster: Waxmann.
- Häußler, P. (1973). Bisherige Ansätze zu disziplinübergreifenden naturwissenschaftlichen Curricula – eine Übersicht. In K. Frey & P. Häußler (Hrsg.), *Integriertes Curriculum Naturwissenschaft: Theoretische Grundlagen und Ansätze* (S. 31–69). Weinheim: Beltz.
- Heering, P., & Höttecke, D. (2011). Was ist Physik? In H. Wiesner, H. Schecker, & M. Hopf (Hrsg.), *Physikdidaktik kompakt* (S. 9–15). Freising: Aulis.
- Heimann, P. (1976a). Didaktik als Theorie und Lehre. In K. Reich & H. Thomas (Hrsg.), *Didaktik als Unterrichtswissenschaft* (S. 142–167). Stuttgart: Klett.
- Heimann, P. (1976b). Didaktische Grundbegriffe, Teil 1: Vortrag vom 07.12.1961. In K. Reich & H. Thomas (Hrsg.), *Didaktik als Unterrichtswissenschaft* (S. 103–121). Stuttgart: Klett.
- Heimann, P., Otto, G., & Schulz, W. (Hrsg.). (1972). *Unterricht – Analyse und Planung* (6. Aufl.). Hannover: Schroedel.
- Helfferich, C. (2011). *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews* (4. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Helfferich, C. (2014). Leitfaden- und Experteninterviews. In N. Baur & J. Blasius (Hrsg.), *Methoden der empirischen Sozialforschung* (S. 559–574). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Helmke, A. (2006). Was wissen wir über guten Unterricht? Über die Notwendigkeit einer Rückbesinnung auf den Unterricht als dem „Kerngeschäft“ der Schule. *Pädagogik*, 58(2), 42–45.
- Helmke, A. (2007). Lernprozesse anregen und steuern – Was wissen wir über Klarheit und Strukturiertheit? *Pädagogik*, 59(5), 44–47.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4. Aufl.). Seelze: Kallmeyer.
- Helmke, A., & Helmke, T. (2004). Videobasierte Unterrichtsreflexion. *Seminar*, 10(4), 48–66.

- Helmke, A., & Schrader, F.-W. (2009). Qualitätsmerkmale „guten Unterrichts“. In S. Hellekamps, W. Plöger & W. Wittenbruch (Hrsg.), *Handbuch für Erziehungswissenschaft, Band II/1: Schule* (S. 701–712). Paderborn: Schöningh.
- Herweg, C. (2008). *Zierorientierung im deutschen und schweizerischen Physikunterricht – eine Videostudie*. Elektronische Dissertation: Christian-Albrechts-Universität Kiel.
- Herzmann, P., & König, J. (2016). *Lehrerberuf und Lehrerbildung*. Bad Heilbrunn: Klinkhardt.
- Herzmann, P., & Proske, M. (2014). Unterrichtsvideografien als Medium der Beobachtung und Reflexion von Unterricht im Lehramtsstudium. Ein Forschungsbericht. *Journal für LehrerInnenbildung*, 14(1), 33–38.
- Hewson, P. W., Beeth, M. E., & Thorley, N. R. (1998). Teaching for conceptual change. In B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education, Part One* (pp. 199–218). Dordrecht: Kluwer Academics Publishers.
- Hofstein, A., & Lunetta, V. N. (1982). The role of laboratory in science teaching: Neglected aspects of research. *Review of Educational Research*, 52, 201–217.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education. Foundations for the twenty-first century. *Studies in Science Education*, 88, 28–54.
- Hopf, C. (1978). Die Pseudo-Exploration – Überlegungen zur Technik qualitativer Interviews in der Sozialforschung. *Zeitschrift für Soziologie*, 7(2), 97–115.
- Hopf, C. (2009). Qualitative Interviews – ein Überblick. In U. Flick, E. von Kardorff, & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (7. Aufl., S. 349–360). Reinbek: Rowohlt.
- Höttecke, D. (2007). Historisch orientierter Physikunterricht. In S. Mikelskis-Seifert & T. Rabe (Hrsg.), *Physik Methodik. Handbuch für die Sekundarstufe I und II* (S. 57–69). Berlin: Cornelsen Scriptor.
- Höttecke, D. (2008). Was ist Naturwissenschaft? Physikunterricht über die Natur der Naturwissenschaften. *Naturwissenschaften im Unterricht – Physik*, 19(103), 4–11.
- Hron, A. (1982). Interview. In G. L. Huber & H. Mandl (Hrsg.), *Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung* (S. 119–140). Weinheim: Beltz.
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288.
- Hugener, I., Rakoczy, K., Pauli, C., & Reusser, K. (2006). Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lernprozesse. In S. Rahm, I. Mammes, & M. Schratz (Hrsg.), *Schulpädagogische Forschung. Unterrichtsforschung. Perspektiven innovativer Ansätze* (S. 41–53). Innsbruck: StudienVerlag.
- Ivankova, N. V., Creswell, J. W., & Stick, S. (2006). Using mixed methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20.
- Jahn, G., Prenzel, M., Stürmer, K., & Seidel, T. (2011). Varianten einer computergestützten Erhebung von Lehrerkompetenzen: Untersuchungen zu Anwendungen des Tools Observer. *Unterrichtswissenschaft*, 39(2), 136–153.
- Jahn, G., Stürmer, K., Seidel, T., & Prenzel, M. (2014). Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden – Eine Scaling-up Studie des Observer-Projekts. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46(4), 171–180.
- Jamil, F., Sabol, T., Hamre, B., & Pianta, R. (2015). Assessing teachers' skills in detecting and identifying effective interactions in the classroom. *The Elementary School Journal*, 115, 407–432.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Towards a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 122–133.
- Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., Kunter, M., & Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal für Mathematik-Didaktik*, 29(2), 83–107.
- Jung, W. (1986). Alltagsvorstellungen und das Lernen von Physik und Chemie. *Naturwissenschaften im Unterricht – Physik/Chemie*, 34, 2–6.
- Justi, R., & Gilbert, J. (2003). Teacher's view on the nature of models. *International Journal of Science Education*, 25, 1369–1386.
- Kalish, C. W., Kim, S., & Young, A. C. (2012). How young children learn from examples: Descriptive and inferential problems. *Cognitive Sciences*, 36(8), 1427–1448.



- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kattmann, U. (2006). Texte. In H. Gropengießer & U. Kattmann (Hrsg.), *Fachdidaktik Biologie. Band 7* (S. 357–366). Köln: Aulis Verlag Deubner.
- Kattmann, U., Duit, R., Gropengießer, H., & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion – Ein theoretischer Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3(3), 3–18.
- Kelava, A., & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 76–102). Berlin: Springer.
- Kelle, U. (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung. Theoretische Grundlagen und methodologische Konzepte* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kelle, U. (2014). Mixed Methods. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 153–166). Wiesbaden: Springer.
- Kerstin, N. B. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement*, 68(5), 845–861.
- Kerstin, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568–589.
- Kind, V. (2009). Pedagogical content knowledge in science education: Perspectives and potential for progress. *Studies in Science Education*, 45(2), 169–204.
- Kircher, E. (1995). *Studien zur Physikdidaktik: Erkenntnis- und wissenschaftstheoretische Grundlagen*. Kiel: IPN.
- Kircher, E. (2015a). Einführung: Was ist Physikdidaktik? In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (3. Aufl., S. 3–14). Berlin: Springer.
- Kircher, E. (2015b). Warum Physikunterricht? In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (3. Aufl., S. 15–72). Berlin: Springer.
- Kircher, E. (2015c). Modellbegriff und Modellbildung in der Physikdidaktik. In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (3. Aufl., S. 783–807). Berlin: Springer.
- Kircher, E. (2015d). Über die Natur der Naturwissenschaften lernen. In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (3. Aufl., S. 809–841). Berlin: Springer.
- Kircher, E., Girwidz, R., & Häußler, P. (Hrsg.). (2015). *Physikdidaktik. Theorie und Praxis* (3. Aufl.). Berlin, Heidelberg: Springer.
- Kirschner, S. B. (2013). *Modellierung und Analyse des Professionswissens von Physiklehrkräften*. Berlin: Logos.
- Klafki, W. (1961). Die didaktischen Prinzipien des Elementaren, Fundamental und Exemplarischen. In A. Blumenthal, J. Guthmann, W. Horney, F. Seilnacht, & K. Stöcker (Hrsg.), *Handbuch für Lehrer* (S. 120–139). Gütersloh: Mohn.
- Klafki, W. (1995). Die bildungstheoretische Didaktik im Rahmen kritisch-konstruktiver Erziehungswissenschaft. Oder: Zur Neufassung der Didaktischen Analyse. In H. Gudjons, R. Teske, & R. Winkel (Hrsg.), *Didaktische Theorien* (8. Aufl., S. 11–27). Hamburg: Bergmann und Helbig.
- Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* (6. Aufl.). Weinheim: Beltz.
- Kleickmann, T. (2012). *Kognitiv aktivieren und inhaltlich strukturieren im naturwissenschaftlichen Sachunterricht. Handreichungen des Programms SINUS an Grundschulen*. Kiel: IPN.
- Kleinknecht, M., Maier, U., Metz, K., & Bohl, T. (2011). Analyse des kognitiven Aufgabenpotentials: Entwicklung und Erprobung eines allgemeindidaktischen Auswertungsmanuals. *Unterrichtswissenschaft*, 39(4), 328–344.
- Klenck, W., & Schneider, S. (2006). Den Unterricht nicht vor dem Ende loben. Plädoyer für die stärkere Beachtung der Schlussphase. *Pädagogik*, 58(7/8), 68–71.
- Klieme, E. (2006). Empirische Unterrichtsforschung: aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 52(6), 765–773.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903.

- Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunkt-programms* (S. 127–146). Münster: Waxmann.
- Klingberg, L. (1978). *Einführung in die Allgemeine Didaktik. Vorlesungen* (4. Aufl.). Berlin: Volk und Wissen.
- Klopfer, L. (1971). Evaluation of learning in science. In B. Bloom (Ed.), *Handbook of formative and summative evaluation of student learning* (pp. 599–641). New York: Mc Graw Hill.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Hrsg.). (2004a). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss (Jahrgangsstufe 10). Beschluss der Kultusministerkonferenz vom 16.12.2004*. Abgerufen von: [http://www.kmk.org/fileadmin/Daten/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Bildungsstandards-Physik-Mittleren-SA.pdf](http://www.kmk.org/fileadmin/Daten/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Physik-Mittleren-SA.pdf)
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Hrsg.). (2004b). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Physik. Beschluss der Kultusministerkonferenz vom 01.12.1989 i. d. F. vom 05.02.2004*. Abgerufen von: [https://www.kmk.org/fileadmin/Daten/veroeffentlichungen\\_beschluesse/1989/1989\\_12\\_01-EPA-Physik.pdf](https://www.kmk.org/fileadmin/Daten/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Physik.pdf)
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Hrsg.). (2013). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. Beschluss der Kultusministerkonferenz vom 16.10.2008 i. d. F. vom 16.05.2013*. Abgerufen von: <https://www.bzl.uni-bonn.de/dokumente/dokumente-1/laendergemeinsame-inhaltliche-anforderungen-fuer-die-fachwissenschaften-und-die-fachdidaktiken-in-der-lehrerbildung>
- Köhler, W. (1969). *The task of Gestalt Psychology*. Princeton, NJ: Princeton University Press.
- König, J. (2010). Lehrerprofessionalität – Konzepte und Ergebnisse der internationalen und deutschen Forschung am Beispiel fachübergreifender, pädagogischer Kompetenzen. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften – Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (S. 40–105). Berlin: DGLS.
- König, J. (2016). Lehrerexpertise und Lehrerkompetenz. In M. Rothland (Hrsg.), *Beruf Lehrer/Lehrerin. Ein Studienbuch* (S. 127–148). Münster: Waxmann.
- König, J., & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5, 3–29.
- König, J., & Seifert, A. (Hrsg.). (2012). *Lehramtsstudierende erwerben professionelles Wissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung*. Münster: Waxmann.
- König, J., Buchholtz, C., & Dohmen, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18(2), 375–404.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, 38, 76–88.
- Kowal, S., & O'Connell, D. C. (2009). Zur Transkription von Gesprächen. In U. Flick, E. von Kardorff, & Steinke, I. (Hrsg.), *Qualitative Forschung. Ein Handbuch* (7. Aufl., S. 437–447). Reinbek: Rowohlt.
- Kramer, C., König, J., Kaiser, G., Ligtvoet, R., & Blömeke, S. (2017). Der Einsatz von Unterrichtsvideos in der universitären Ausbildung: Zur Wirksamkeit video- und transkriptgestützter Seminare zur Klassenführung auf pädagogisches Wissen und situationsspezifische Fähigkeiten angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 20(Supplement 1), 137–164.
- Krammer, K., & Hugener, I. (2014). Förderung der Analysekompetenz angehender Lehrpersonen anhand von eigenen und fremden Unterrichtsvideos. *Journal für LehrerInnenbildung*, 14(1), 25–32.
- Krammer, K., & Reusser, K. (2005). Unterrichtsvideos als Medium der Aus- und Weiterbildung von Lehrpersonen. *Beiträge zur Lehrerbildung*, 23(1), 35–50.
- Krauss, S. (2011). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz, M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 171–191). Münster: Waxmann.
- Krauss, S., & Bruckmaier, G. (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., S. 240–261). Münster: Waxmann.
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., & Neubrand, M. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Die Bildungsqualität von Schule: Lehrerprofessionalisierung*,

- Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 31–53). Münster: Waxmann.
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M., & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3/4), 223–258.
- Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., Besser, M., & Elsner, J. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 135–161). Münster: Waxmann.
- Krems, J. (1996). Expertise und Flexibilität. In H. Gruber & A. Ziegler (Hrsg.), *Expertiseforschung. Theoretische und methodische Grundlagen* (S. 80–91). Opladen: Westdeutscher Verlag.
- Kron, F. W. (2008). *Grundwissen Didaktik* (5. Aufl.). München: Reinhardt.
- Krull, E., Oras, K., & Pikksaar, E. (2010). Promoting student teachers' lesson analysis and observation skills by using Gangé's model of an instructional unit. *Journal of Education for Teaching*, 36(2), 197–210.
- Kuckartz, U. (2014). *Mixed Methods. Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: Springer VS.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (3. Aufl.). Weinheim: Beltz Juventa.
- Kuhn, Th. (1979). *Die Struktur wissenschaftlicher Revolutionen* (4. Aufl.). Frankfurt am Main: Suhrkamp.
- Kulgemeyer, C., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., Riese, J., ... Walzer, M. (2012). ProfiLe-P – Professionswissen in der Lehramtsausbildung Physik. Vorstellung eines Forschungsverbundes. In V. Nordmeier & H. Grötzebach (Hrsg.), *PhyDid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung 2012 in Mainz* (S. 1–6). Abgerufen von: <http://www.phydid.de/index.php/phydid-b/article/view/380/500>
- Kunter, M., & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. Gebauer, & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 9–31). Münster: Waxmann.
- Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts*. Paderborn: Schöningh.
- Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: A multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV Project* (pp. 97–124). New York: Springer.
- Kunter, M., Klusmann, U., & Baumert, J. (2009). Professionelle Kompetenz von Mathematiklehrkräften: Das COACTIV-Modell. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus, & R. Mulder (Hrsg.), *Lehrerprofessionalität – Bedingungen, Genese, Wirkungen und ihre Messung* (S. 153–165). Weinheim: Beltz.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., Jordan, A., & Neubrand, M. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler. *Zeitschrift für Erziehungswissenschaft*, 8(4), 502–520.
- Kyriakides, L., Christoforou, C., & Charalambous, Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152.
- Labudde, P. (1997). Selbständig lernen. Eine Chance für den Physikunterricht. *Unterricht Physik* 37, 4–9.
- Labudde, P. (2006). Physikalische Bildung mit Zukunft: Zehn Thesen. In R. Girwidz, M. Gläser-Zikuda, M. Lauckenmann, & T. Rubitzko (Hrsg.), *Lernen im Physikunterricht. Festschrift für Christoph von Rhöneck* (S. 251–262). Hamburg: Kovač.
- Lamnek, S. (2010). *Qualitative Sozialforschung* (5. Aufl.). Weinheim: Beltz.
- Langer, A. (2010). Transkribieren – Grundlagen und Regeln. In B. Frieberthäuser, A. Langer, & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl., S. 515–526). Weinheim: Juventa.
- Latcheva, R. (2009). Cognitive interviewing and factor-analytic techniques: A mixed method approach to validity of survey items measuring national identity. *Quality and Quantity*, 45(6), 1175–1199.
- Lauf, E. (2001). .96 nach Holsti. Zur Reliabilität von Inhaltsanalysen und deren Darstellung in kommunikationswissenschaftlichen Fachzeitschriften. *Publizistik*, 1, 57–68.

- Lederman, N. G., Gess-Newsome, J., & Latz, M. S. (1994). The nature and development of preservice science teachers' conceptions of subject matter and pedagogy. *Journal of Research in Science Teaching*, 31(2), 129–146.
- Leech, N. L., Dellinger, A. B., Brannagan, K. B., & Tanaka, H. (2010). Evaluating mixed research studies: A mixed methods approach. *Journal of Mixed Methods Research*, 4(1), 17–31.
- Leinhardt, G., & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), 75–95.
- Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, 77(3), 247–271.
- Leisner, A. (2005). *Entwicklung von Modellkompetenz im Physikunterricht*. Berlin: Logos.
- Leuders, T., & Holzäpfel, L. (2011). Kognitive Aktivierung im Mathematikunterricht. *Unterrichtswissenschaft*, 39(3), 213–230.
- Lienert, G. A. (1975). *Verteilungsfreie Methoden in der Biostatistik, Tafelband*. Meisenheim am Glan: Hain.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz, Psychologie Verlags Union.
- Lindmeier, A. (2011). *Modeling and Measuring Knowledge and Competencies of Teachers. A Threefold Domain-Specific Structure Model for Mathematics*. Münster: Waxmann.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf* (51. Beiheft der Zeitschrift für Pädagogik, S. 47–70). Weinheim: Beltz.
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (2. Aufl., S. 69–105). Berlin: Springer.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19(6), 527–537.
- Lissmann, U. (2008). *Inhaltsanalyse von Texten: ein Lehrbuch zur computerunterstützten und konventionellen Inhaltsanalyse* (3. Aufl.) Landau: Verlag Empirische Pädagogik.
- Livingston, C., & Borko, H. (1990). High school mathematics review lessons: Expert-Novice distinctions. *Journal of Research in Mathematics Education*, 21(5), 372–387.
- Lunetta, V. N. (1998). The school science laboratory: Historical perspectives and contexts for contemporary teaching. In B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education, Part One* (pp. 249–262). Dordrecht: Kluwer Academic Publishers.
- Luyt, R. (2011). A framework for mixing methods in quantitative measurement development validation, and revision: A case study. *Journal of Mixed Methods Research*, 6(4), 294–316.
- Magnusson, S., Krajcik, J. S., & Borko, H. (1999). Secondary teachers' knowledge and beliefs about subject matter and their impact on instruction. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Dordrecht: Kluwer.
- Maier, U., Bohl, T., Kleinknecht, M., & Metz, K. (2013). Allgemeindidaktische Kriterien für die Analyse von Aufgaben. In M. Kleinknecht, T. Bohl, U. Maier, & K. Metz (Hrsg.), *Lern- und Leistungsaufgaben im Unterricht. Fächerübergreifende Kriterien zur Auswahl und Analyse* (S. 9–46). Bad Heilbrunn: Klinkhardt.
- Mandl, H., & Kopp, B. (2005). Situated learning: theories and models. In P. Nentwig & D. Waddington (Eds.), *Context based learning in science* (pp. 15–34). Münster: Waxmann.
- Marsch, S., Hartwig, C., & Krüger, D. (2009). Lehren und Lernen im Biologieunterricht. Ein Kategoriensystem zur Beurteilung konstruktivistisch orientierter Lernumgebungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 109–130.
- Mayer, R. E. (2002). Rote versus Meaningful Learning. *Theory into Practice*, 41(4), 226–232.
- Mayring, P. (1995). Qualitative Inhaltsanalyse. In U. Flick, E. von Kardorff, H. Keupp, L. von Rosenstiel, & S. Wolff (Hrsg.), *Handbuch qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen* (2. Aufl., S. 209–213). Weinheim: Beltz, Psychologie Verlags Union.
- Mayring, P. (2000). Qualitative Inhaltsanalyse [28 Absätze]. *Forum Qualitative Sozialforschung/Forum Qualitative Research*, 1(2), Art. 20. Abgerufen von: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204>
- Mayring, P. (2001). Kombination und Integration qualitativer und quantitativer Analyse [31 Absätze]. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research*, 2(1), Art. 6. Abgerufen von: <http://nbn-resolving.de/urn:nbn:de:0114-fqs010162>
- Mayring, P. (2002). *Einführung in die qualitative Sozialforschung* (5. Aufl.). Weinheim: Beltz Verlag.

- Mayring, P. (2008). Neuere Entwicklungen in der qualitativen Forschung und der Qualitativen Inhaltsanalyse. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (2. Aufl., S. 7–19). Weinheim: Beltz.
- Mayring, P. (2009). Qualitative Inhaltsanalyse. In U. Flick, E. von Kardorff, & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (7. Aufl., S. 468–475). Reinbek: Rowohlt.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 601–613). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mayring, P. (2012). Qualitative Inhaltsanalyse – ein Beispiel für Mixed Methods. In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner, & S. Ziegelbauer (Hrsg.), *Mixed Methods in der empirischen Bildungsforschung* (S. 27–36). Münster: Waxmann.
- Mayring, P. (2014). Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt. Abgerufen von: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173>
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- Mayring, P., & Brunner, E. (2010). Qualitative Inhaltsanalyse. In B. Frieberthäuser, A. Langer, & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl., S. 323–333). Weinheim: Juventa.
- Mayring, P., & Fenzl, T. (2014). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 543–556). Wiesbaden: Springer.
- Meisert, A. (2008). Vom Modellwissen zum Modellverständnis – Elemente einer umfassenden Modellkompetenz und deren Fundierung durch lernseitige Kriterien zur Klassifikation von Modellen. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 243–261.
- Merton, R. K., & Kendall, P. L. (1993). Das fokussierte Interview. In C. Hopf & E. Weingarten (Hrsg.), *Qualitative Sozialforschung* (3. Aufl., S. 171–204). Stuttgart: Klett-Cotta.
- Merton, R. K., Fiske, M., & Kendall, P. L. (1956). *The Focused Interview. A Manual of Problems and Procedures*. Glencoe, IL: Free Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement on Education.
- Messick, S. (1990). *Validity of test interpretation and use* (Report No. ETS-RR. 90.11). Princeton, NJ: Educational Testing Service.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Meuser, M., & Nagel, U. (1991). Experteninterviews – vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In D. Garz & K. Kraimer (Hrsg.), *Qualitativ-empirische Sozialforschung. Konzepte, Methoden, Analysen* (S. 441–471). Opladen: Westdeutscher Verlag.
- Meuser, M., & Nagel, U. (2009). Experteninterview und der Wandel der Wissensproduktion. In A. Bogner, B. Littig, & W. Menz (Hrsg.), *Experteninterviews. Theorien, Methoden, Anwendungsfelder* (3. Aufl., S. 35–60). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Meuser, M., & Nagel, U. (2010). Experteninterviews – wissenssoziologische Voraussetzungen und methodische Durchführung. In B. Frieberthäuser, A. Langer, & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl., S. 457–471). Weinheim: Juventa.
- Meyer, H. (1987). *Unterrichtsmethoden. Band II Praxisband*. Frankfurt am Main: Cornelsen Scriptor.
- Meyer, H. (2003). Zehn Merkmale guten Unterrichts. Empirische Befunde und didaktische Ratschläge. *Pädagogik*, 55(10), 36–43.
- Meyer, H. (2013). *Was ist guter Unterricht?* (9. Aufl.). Berlin: Cornelsen Scriptor.
- Mikelskis, H. (2007). Exemplarisch – genetisch – sokratisch. In S. Mikelskis-Seifert & T. Rabe (Hrsg.), *Physik Methodik. Handbuch für die Sekundarstufe I und II* (S. 44–56). Berlin: Cornelsen Scriptor.
- Mikelskis-Seifert, S., & Duit, R. (2010). *Physik im Kontext – Konzepte, Ideen, Materialien für einen effizienten Physikunterricht*. Seelze: Friedrich Verlag.
- Mikelskis-Seifert, S., & Fischler, H. (2003). Die Bedeutung des Denkens in Modellen bei der Entwicklung von Teilchenvorstellungen – Empirische Untersuchung zur Wirksamkeit der Unterrichtskonzeption. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 89–103.
- Mikelskis-Seifert, S., & Leisner, A. (2003). Das Denken in Modellen fördern. Ein Unterrichtsbeispiel zur Entwicklung von Teilchenvorstellungen. *Unterricht Physik*, 14, 32–34.
- Mikelskis-Seifert, S., & Leisner-Bodenthin, A. (2007). Unterrichtsansätze zu den Methoden physikalischer Erkenntnisgewinnung. In S. Mikelskis-Seifert & T. Rabe, *Physik Methodik. Handbuch für die Sekundarstufe I und II* (S. 15–43). Berlin: Cornelsen Scriptor.



- Mikelskis-Seifert, S., & Rabe, T. (Hrsg.). (2007). *Physik Methodik. Handbuch für die Sekundarstufe I und II*. Berlin: Cornelsen Scriptor.
- Misoch, S. (2015). *Qualitative Interviews*. Berlin: de Gruyter.
- Morell, L., & Tan, R. J. B. (2009). Validating for use and interpretation. A mixed methods contribution illustrated. *Journal of Mixed Methods Research*, 3(3), 242–264.
- Morse, J. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(2), 120–123.
- Moosbrugger, H., & Schermelleh-Engel (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 325–343). Heidelberg: Springer.
- Muckenfuß, H. (1995). *Lernen im sinnstiftenden Kontext*. Berlin: Cornelsen.
- Muckenfuß, H. (2004). Themen und Kontexte als Strukturelemente des naturwissenschaftlichen Unterrichts. Zu den Schwierigkeiten, systematisches Physiklernen zu organisieren. *Physik und Didaktik in Schule und Hochschule*, 2/3 2004, 57–66.
- Muijs, D., & Reynolds, D. (2011). *Effective teaching: evidence and practice* (3<sup>rd</sup> Ed.). London: Sage.
- Mulder, R. H., & Gruber, H. (2011). Die Lehrperson im Lichte von Professions-, Kompetenz- und Expertiseforschung – die drei Seiten einer Medaille. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Bildungsforschung. Traditionslinien und Perspektiven* (S. 427–438). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Müller, R., Wodzinski, R., & Hopf, M. (Hrsg.). (2004). *Schülervorstellungen in der Physik*. Köln: Aulis.
- Munby, H., Russell, T., & Martin, A. K. (2002). Teacher's knowledge and how it develops. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 877–904). Washington: American Educational Research Association.
- Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2011). Aufgaben im COACTIV-Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 115–132). Münster: Waxmann.
- Nilsson, P., & Loughran, J. (2012). Exploring the development of pre-service science elementary teachers' pedagogical content knowledge. *Journal of Science Teacher Education*, 23(7), 699–721.
- Ohle, A., Fischer, H. E., & Kauertz, A. (2011). Der Einfluss physikalischen Fachwissens von Primarstufenlehrkräften auf Unterrichtsgestaltung und Schülerleistung. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 357–389.
- Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed methods research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research*, 4(1), 56–78.
- Oser, F., & Baeriswyl, F. J. (2002). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4<sup>th</sup> Ed., pp. 1031–1065). Washington: American Educational Research Association.
- Oser, F., & Patry, J.-L. (1994). Sichtstruktur und Basismodelle des Unterrichts: Über den Zusammenhang von Lehren und Lernen unter dem Gesichtspunkt psychologischer Lernverläufe. In R. Olechowski & B. Rollet (Hrsg.), *Theorie und Praxis – Aspekte empirisch-pädagogischer Forschung – quantitative und qualitative Methoden* (S. 138–146). Frankfurt am Main: Lang.
- Oser, F., & Renold, U. (2005). Kompetenzen von Lehrpersonen – Über das Auffinden von Standards und ihre Messung. *Zeitschrift für Erziehungswissenschaft*, 8(4), 119–140.
- Palmer, D. J., Stough, L. M., Burdinski, T. K., & Conzaes, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Educational Psychologist*, 40(1), 13–25.
- Panasuk, R. M., & Sullivan, M. (1998). Need for lesson analysis in effective lesson planning. *Education*, 118, 330–344.
- Paradies, L., & Linser, H. J. (2010). *Differenzieren im Unterricht* (5. Aufl.). Berlin: Cornelsen Scriptor.
- Parchmann, I., Ralle, B., & Demuth, R. (2000). Chemie im Kontext – eine Konzeption zum Aufbau und zur Aktivierung fachsystematischer Strukturen in lebensweltlichen Kontexten. *MNU*, 53, 132–136.
- Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49, 261–284.
- Park, S., & Oliver, S. J. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284.

- Park, S., Jang, J.-Y., Chen, Y.-C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching? Evidence from an empirical study. *Research in Science Education*, 41(2), 245–260.
- Pauli, C., & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52(6), 774–798.
- Peterson, P. L., & Comeaux, M. A. (1987). Teachers' schemata for classroom events: The mental scaffolding of teachers' thinking during classroom instruction. *Teaching and Teacher Education*, 3(4), 319–331.
- Peterßen, W. H. (2000). *Handbuch Unterrichtsplanung. Grundfragen, Modelle, Stufen, Dimensionen* (9. Aufl.). München: Oldenbourg.
- Planck, M. (1948). *Wissenschaftliche Selbstbiographie. Mit einem Bildnis und der von Max von Laue gehaltenen Traueransprache*. Leipzig: Barth.
- Plöger, W. (1983). *Forschender Physikunterricht. Grundlegung und Unterrichtsentwürfe*. Ansbach: Prögel.
- Plöger, W. (2001). Genetisches Lehren und Lernen nach Martin Wagenschein. In M. Brenk & U. Kurth (Hrsg.), *SCHULE erLEBEN. Festschrift für Wilhelm Wittenbruch* (S. 149–164). Frankfurt am Main: Lang.
- Plöger, W. (2006). Naturwissenschaftlicher Unterricht unter dem Anspruch von Allgemeinbildung. *Vierteljahresschrift für Wissenschaftliche Pädagogik*, 82(1), 19–31.
- Plöger, W. (2008). *Unterrichtsplanung. Ein Lehr- und Arbeitsbuch für Studium- und Seminar*. Köln: Kölner Studien Verlag.
- Plöger, W. (2011). Lernen – Ein blinder Fleck in Unterrichtstheorien? In W. Meseth, M. Proske & F.-O. Radtke (Hrsg.), *Unterrichtstheorien in Forschung und Lehre* (S. 85–102). Bad Heilbrunn: Klinkhardt.
- Plöger, W., & Scholl, D. (2012). Analysekompetenz von Lehrpersonen – Diagnose und Entwicklung. In T. Ha-scher und G.H. Neuweg (Hrsg.), *Forschung zur (Wirksamkeit der) Lehrer/innen/bildung* (S. 239–257). Wien: LIT.
- Plöger, W., & Scholl, D. (2013). Didaktisch-methodische Kompetenzen und ihre Bedeutung für die Analyse von Unterricht. In K. Zierer, u.a. (Hrsg.), *Jahrbuch für Allgemeine Didaktik 2013. Thementeil: Neuere Ansätze in der Allgemeinen Didaktik* (S. 133–146). Baltmannsweiler: Schneider Verlag Hohengehren.
- Plöger, W., & Scholl, D. (2014a). Analysekompetenz von Lehrpersonen – Modellierung und Messung. *Zeitschrift für Erziehungswissenschaft*, 17(1), 85–112.
- Plöger, W., & Scholl, D. (2014b). Wie gut können Lehrpersonen Unterricht analysieren? Ergebnisse eines empirischen Forschungsprojektes. *Pädagogische Rundschau*, 68(3), 273–287.
- Plöger, W., & Scholl, D. (2014c). Analysekompetenz angehender und praktizierender Lehrpersonen. In A. Bresges, B. Dilger, T. Hennemann, J. König, H. Lindner, A. Rohde, & D. Schmeinc (Hrsg.), *Kompetenzen diskursiv. Terminologische, exemplarische und strukturelle Klärungen in der LehrerInnenbildung* (S.137–156). Münster: Waxmann.
- Plöger, W., & Scholl, D. (2015). *Pädagogisches Wissen – eine notwendige Voraussetzung für die Analyse von Unterricht*. Unveröffentlichtes Manuskript, Institut für Allgemeine Didaktik und Schulforschung, Universität zu Köln, Köln, Deutschland.
- Plöger, W., & Scholl, D. (2016). Die Entwicklung didaktisch-methodischer Kompetenzen zur Verbesserung der Analyse von Unterricht. In A. Wegner (Hrsg.), *Allgemeine Didaktik: Praxis, Positionen, Perspektiven* (S. 181–198). Opladen: Budrich.
- Plöger, W., Scholl, D., & Seifert, A. (2015). Analysekompetenz – ein zweidimensionales Konstrukt. *Unterrichtswissenschaft*, 42(2), 166–184.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
- Pospeschill, M. (2006). *Statistische Grundlagen. Strukturen, Grundlagen, Anwendungen in Psychologie und Sozialwissenschaften*. München: Elsevier.
- Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation*. München: Reinhardt.
- Prenzel, M., Carstensen, C. H., Senkbeil, M., Ostermeier, C., & Seidel, T. (2005). Wie schneiden SINUS-Schulen bei PISA ab? Ergebnisse der Evaluation eines Modellversuchsprogramms. *Zeitschrift für Erziehungswissenschaft*, 8(4), 540–561.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P., & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 191–248). Opladen: Leske + Budrich.
- Przyborski, A., & Wohlrab-Sahr, M. (2014). *Qualitative Sozialforschung. Ein Arbeitsbuch* (4. Aufl.). München: Oldenbourg.
- Reinders, U. (2015). Interview. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (2. Aufl., S. 93–107). Wiesbaden: Springer.



- Reiss, K. (2002). Beweisen, Begründen, Argumentieren. Wege zu einem diskursiven Mathematikunterricht. In W. Peschek (Hrsg.), *Beiträge zum Mathematikunterricht 2002*. Vorträge auf der 36. Tagung für Didaktik der Mathematik vom 25. Februar bis 1. März 2002 in Klagenfurt (S. 39–46). Hildesheim: Franzbecker.
- Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird. *Psychologische Rundschau*, 47(2), 541–562.
- Renkl, A. (1998). Träges Wissen. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 514–516). Weinheim: Psychologie Verlags Union.
- Renkl, A. (2001). Üben. In W. Einsiedler, M. Götz, H. Hacker, J. Kahlert, R. W. Keck, & U. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (S. 361–367). Bad Heilbrunn: Klinkhardt.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. *Educational Psychologist*, 38(1), 15–22.
- Renkl, A., & Nückles, M. (2006). Lernstrategien der externen Visualisierungen. In H. Mandl & H. F. Friedrich (Hrsg.), *Handbuch Lernstrategien* (S. 135–147). Göttingen: Hogrefe.
- Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review*, 21(1), 67–78.
- Renkl, A., Atkinson, R. K., Maier, U. H., & Stanley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education*, 70(4), 293–315.
- Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out-examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, 23(1), 90–108.
- Reusser, K., & Pauli, C. (2003). *Mathematikunterricht in der Schweiz und in weiteren sechs Ländern. Bericht über die Ergebnisse einer internationalen und schweizerischen Video-Unterrichtsstudie*. CD-Rom. Zürich: Universität Zürich.
- Reyer, T. (2004). *Oberflächenmerkmale und Tiefenstrukturen im Unterricht. Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*. Berlin: Logos.
- Reynolds, A. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research*, 62(1), 1–35.
- Riese, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften*. Berlin: Logos.
- Riese, J., & Reinhold, P. (2009). Fachbezogene Kompetenzmessung und Kompetenzentwicklung bei Lehramtsstudierenden der Physik im Vergleich verschiedener Studiengänge. *Lehrerbildung auf dem Prüfstand*, 2(1), 104–125.
- Riese, J., & Reinhold, P. (2010). Empirische Erkenntnisse zur Struktur professioneller Handlungskompetenz von angehenden Physiklehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 167–187.
- Riese, J., & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen. Empirische Hinweise für eine Verbesserung des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft*, 15(1), 111–143.
- Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H., Gramzow, Y., ... Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (61. Beiheft der Zeitschrift für Pädagogik, S. 55–79). Weinheim: Beltz Juventa.
- Rincke, K. (2012). Alltagssprache, Fachsprache und ihre besonderen Bedeutungen für das Lernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 235–260.
- Rincke, K., & Leisen, J. (2015). Sprache im Physikunterricht. In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (3. Aufl., S. 635–680). Berlin: Springer.
- Rollet, B. (1999). Auf dem Weg zu einer Fehlerkultur. Anmerkungen zur Fehlertheorie von Fritz Oser. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlass des 60. Geburtstages von Fritz Oser* (S. 71–87). Opladen: Leske + Budrich.
- Rosaen, C. L., Carlisle, J. F., Mihocko, E., Melnick, A., & Johnson, J. (2013). Teachers learning from analysis of other teachers' reading lessons. *Teaching and Teacher Education*, 35, 170–184.
- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research of teaching* (3<sup>rd</sup> Ed., pp. 85–107). New York: Macmillan.
- Rosbach, H. G., & Wellenreuther, M. (2002). Empirische Forschungen zur Wirksamkeit von Methoden der Leistungsdifferenzierung in der Grundschule. In F. Heinzel & A. Prengel (Hrsg.), *Heterogenität, Integration und Differenzierung in der Primarstufe* (S. 44–57). Opladen: Leske + Budrich.
- Rost, D. H. (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung* (3. Aufl.). Bad Heilbrunn: Klinkhardt.

- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Huber.
- Rubin, A., & Babbie, E. R. (2015). *Essential research methods for social work* (4. Aufl.). Belmont, CA: Brooks/Cole Cengage Learning.
- Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal*, 28(1), 63–88.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Schermelleh-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 119–141). Heidelberg: Springer.
- Scheuch, E. K. (1973). Das Interview in der Sozialforschung. In R. König (Hrsg.), *Handbuch der empirischen Sozialforschung. Band 2 Grundlegende Methoden und Techniken der empirischen Sozialforschung. Erster Teil* (3. Aufl., S. 66–190). Stuttgart: Enke.
- Schittko, K. (1984). *Differenzierung in Schule und Unterricht: Ziele – Konzepte – Beispiele*. München: Ehrenwirth.
- Schmidt, C. (1997). „Am Material“: Auswertungstechniken für Leitfadeninterviews. In B. Friebertshäuser & A. Pregel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (S. 544–568). Weinheim: Juventa.
- Schmidt-Wulffen, W.-D. (1989). Diskursive Problemerkennung statt linearer Wissensvermittlung – am Beispiel des Sahelnomadismus. *Geographica Helvetica*, 44, 22–28.
- Schmotz, C., Felbrich, A., & Kaiser, G. (2010). Überzeugungen angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich. In S. Blömeke, G. Kaiser, & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich* (S. 279–305). Münster: Waxmann.
- Schreier, M. (2010). Fallauswahl. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 238–251). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schreier, M. (2011). Qualitative Stichprobenkonzepte. In G. Naderer & E. Balzer (Hrsg.), *Qualitative Marktforschung in Theorie und Praxis* (S. 243–256). Wiesbaden: Gabler.
- Schreier, M. (2013a). Qualitative Forschungsmethoden. In W. Hussy, M. Schreier, & G. Echterhoff (Hrsg.), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (2. Aufl., S. 189–221). Berlin: Springer.
- Schreier, M. (2013b). Qualitative Erhebungsmethoden. In W. Hussy, M. Schreier, & G. Echterhoff (Hrsg.), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (2. Aufl., S. 222–244). Berlin: Springer.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten [59 Absätze]. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research*, 15(1), Art. 18. Abgerufen von: <http://nbn-resolving.de/urn:nbn:de:0114-fqs1401185>
- Schreier, M., & Echterhoff, G. (2013). Mixed-Methods-Designs. In W. Hussy, M. Schreier, & G. Echterhoff (Hrsg.), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (2. Aufl., S. 298–310). Berlin: Springer.
- Schreier, M., & Odag, Ö. (2010). Mixed Methods. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 263–277). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schröder, H. (2002). *Lernen – Lehren – Unterricht. Lernpsychologische und didaktische Grundlagen* (2. Aufl.). München: Oldenbourg.
- Schulz, W. (1972). Unterricht – Analyse und Planung. In P. Heimann, G. Otto, & W. Schulz (Hrsg.), *Unterricht. Analyse und Planung* (6. Aufl., S. 13–47). Hannover: Schroedel.
- Schulz, W. (2006). Die lehrtheoretische Didaktik. Oder: Didaktisches Handeln im Schulfeld. Modellskizze einer professionellen Tätigkeit. In H. Gudjons & R. Winkel (Hrsg.), *Didaktische Theorien* (12. Aufl.). Hamburg: Bergmann und Helbig.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775.
- Sedlmeier, P., & Renkewitz, F. (2013). *Forschungsmethoden und Statistik. Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (2. Aufl.). München: Pearson.
- Seidel, T. (2003a). *Lehr-Lernskripts im Unterricht. Freiräume und Einschränkungen für kognitive und motivationale Lernprozesse – eine Videostudie im Physikunterricht*. Münster: Waxmann.
- Seidel, T. (2003b). Sichtstrukturen – Organisation unterrichtlicher Aktivitäten. In T. Seidel, M. Prenzel, R. Duit, & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 113–127). Kiel: IPN.

- Seidel, T. (2014). Lehrerhandeln im Unterricht. In E. Terhart, H. Bennewitz, & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrberuf* (2. Aufl., S. 781–806). Münster: Waxmann.
- Seidel, T., & Prenzel, M. (2004). Muster unterrichtlicher Aktivitäten im Physikunterricht. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 177–194). Münster: Waxmann.
- Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, 16(3), 228–240.
- Seidel, T., & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (8. Sonderheft der Zeitschrift für Erziehungswissenschaft, S. 201–216). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Seidel, T., & Shavelson, R. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51(4), 739–771.
- Seidel, T., & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaft*, 20(Supplement 1), 1–21.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In E. Klieme, D. Leutner, & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (56. Beiheft der Zeitschrift für Pädagogik, S. 296–306) Weinheim: Beltz.
- Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education*, 27(2), 259–267.
- Seidel, T., Prenzel, M., Rimmele, R., Schwindt, K., Kobarg, M., Herweg, C., & Dalehefte, I. M. (2006). Unterrichtsmuster und ihre Wirkungen. Eine Videostudie im Physikunterricht. In M. Prenzel & L. A. Näcke (Hrsg.), *Untersuchungen zur Qualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 99–123). Münster: Waxmann.
- Seidel, T., Prenzel, M., Rimmele, R., Dalehefte, I. M., Herweg, C., Kobarg, M., & Schwindt, K. (2006). Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie. *Zeitschrift für Pädagogik*, 52(6), 799–821.
- Seiler, T. (1973). Kognitive Strukturen und kognitive Persönlichkeitstheorien. In T. B. Seiler (Hrsg.), *Kognitive Strukturiertheit. Theorien, Analysen, Befunde* (S. 9–27). Stuttgart: Kohlhammer.
- Shavelson, R. J. (1973). What is the basic teaching skill? *Journal of Teacher Education*, 24(2), 144–151.
- Shavelson, R. J. (2013). On an approach to testing and modeling competence. *Educational Psychologist*, 48(2), 73–86.
- Shavelson, R. J., & Stern, O. (1981). Research on teacher's pedagogical thoughts, judgements, decisions, and behavior. *Review of Educational Research*, 51(4), 455–498.
- Shuell, T. J. (1990). Phases of meaningful learning. *Review of Educational Research*, 60(4), 531–547.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Research*, 57(1), 1–22.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21(2), 141–157.
- Sneed, J. (1971). *The Logical Structure of Mathematical Physics*. Dordrecht: Reidel.
- Sommer, W. (2005). *Zur phänomenologischen Beschreibung der Beugung im Konzept optischer Wege*. Berlin: Logos.
- Spychiger, M., Oser, F., Hascher, T., & Mahler, F. (1999). Entwicklung einer Fehlerkultur in der Schule. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlass des 60. Geburtstages von Fritz Oser* (S. 43–70). Opladen: Leske + Budrich.
- Stadler, M. (2009). Modul 2: Naturwissenschaftliches Arbeiten. In M. Prenzel, A. Friedrich, & A. Stadler (Hrsg.), *Von SINUS lernen – Wie Unterrichtsentwicklung gelingt* (S. 23–26). Seelze-Velber: Kallmeyer/Klett.
- Stamann, Ch., Janssen, M., & Schreier, M. (2016). Qualitative Inhaltsanalyse – Versuch einer Begriffsbestimmung und Systematisierung [24 Absätze]. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research*, 17(3), Art. 16. Abgerufen von: <http://nbn-resolving.de/urn:nbn:de:0114-fqs1603166>
- Stegmüller, W. (1979). *Hauptströmungen der Gegenwartsphilosophie. Band II* (6. Aufl.). Stuttgart: Kröner.

- Steigleder, S. (2008). *Die strukturierende qualitative Inhaltsanalyse im Praxistest. Eine konstruktiv kritische Studie zur Auswertungsmethodik von Philipp Mayring*. Marburg: Tectum.
- Steinke, I. (1999). *Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Weinheim: Juventa.
- Steinke, I. (2009). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardorff, & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (7. Aufl., S. 319–331). Reinbek: Rowohlt.
- Strahan, D. B. (1989). How experienced and novice teachers frame their views of instruction: An analysis of semantic ordered trees. *Teaching and Teacher Education*, 5(1), 53–67.
- Strauss, A. L., & Corbin, J. M. (1996). *Grounded Theory: Grundlagen qualitativer Sozialforschung*. Weinheim: Beltz.
- Strobl, C. (2010). *Das Raschmodell. Eine verständliche Einführung für Studium und Praxis*. München: Hampp.
- Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates. Approaches to validating the Observer Extended Tool. *Zeitschrift für Psychologie*, 223(1), 54–63.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology. Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (2003a). Major Issues and Controversies in the Use of Mixed Methods in Social and Behavioral Research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social & Behavioral Research* (pp. 3–50). Thousand Oaks: Sage.
- Tashakkori, A., & Teddlie, C. (2003b). The Past and Future of Mixed Methods Research: From Data Triangulation to Mixed Model Designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social & Behavioral Research* (pp. 671–700). Thousand Oaks: Sage.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., Thillmann, H., & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7–28.
- Tergan, S.-O. (1986). *Modelle der Wissensrepräsentationen als Grundlage qualitativer Wissensdiagnostik*. Opladen: Westdeutscher Verlag.
- Terhart, E. (2002). *Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz* (Universität Münster, ZKL-Texte Nr. 24). Münster: Universität Münster.
- Terhart, E. (2007). Erfassung und Beurteilung der beruflichen Kompetenzen von Lehrkräften. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 37–62). Münster: Waxmann.
- Terhart, E. (2010). Guter Unterricht: Die Perspektiven der allgemeinen Didaktik und der empirischen Unterrichtsforschung. In C. Fischer & R. Schilmöller (Hrsg.), *Was ist guter Unterricht? Qualitätskriterien auf dem Prüfstand* (S. 39–52). Münster: Waxmann.
- Terhart, E. (2016). Zur Relevanz der aktuellen Unterrichtsforschung für die Lehrerbildung. In N. McElvany, W. Bos, H. G. Holtappels, M. Gebauer, & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 171–176). Münster: Waxmann.
- Tesch, M. (2005). *Das Experiment im Physikunterricht: Didaktische Konzepte und Ergebnisse einer Videostudie*. Berlin: Logos.
- Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 51–69.
- Theilmann, F. (2008a). Physikalisches Verstehen als fachbezogene Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 85–97.
- Theilmann, F. (2008b). Der Gedanke im Kontext. Zur Charakteristik einer erscheinungsorientierten Physik. *Der Mathematische und Naturwissenschaftliche Unterricht*, 61, 236–241.
- Tuldozicki, G., Blömeke, S., & Hertz, B. (2009). *Gestaltung von Unterricht* (2. Aufl.). Bad Heilbrunn: Klinkhardt.
- Vielhaber, C. (2001). Diskursiver Unterricht. In W. Sitte & H. Wohlschlägl (Hrsg.), *Beiträge zur Didaktik des „Geographie und Wirtschaftskunde“-Unterrichts* (S. 90–100). Wien: Universität Wien, Institut für Geographie und Regionalforschung.
- Vogelsang, C., & Reinhold, P. (2013). Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 103–128.
- Vollstädt, W. (1997). Differenzierung im Unterricht. Oder: Prüfungswissen kompakt. *Pädagogik*, 49(12), 36–40.

- Vosniadou, S. (2001). *How children learn*. Educational practices series 7. Brüssel: International Academy of Education. Abgerufen von: <http://www.ibe.unesco.org>
- Voss, T., & Kunter, M. (2011). Pädagogisch-psychologisches Wissen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 193–214). Münster: Waxmann.
- Voss, T., Kleickmann, T., Kunter, M., & A. Hachfeld (2011). Überzeugungen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 235–257). Münster: Waxmann.
- Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18(2), 187–223.
- Voss, T., Kunter, M., Seiz, J., Hoehne, V., & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik*, 60(2), 184–201.
- Wagenschein, M. (1965). *Ursprüngliches Verstehen und exaktes Denken*. Stuttgart: Klett.
- Wagenschein, M. (1968). *Verstehen lernen. Genetisch – sokratisch – exemplarisch*. Weinheim: Beltz.
- Wagenschein, M. (1975). *Natur physikalisch gesehen. Didaktische Beiträge zum Vorrang des Verstehens*. Braunschweig: Westermann.
- Walberg, H. J., & Paik, S. J. (2000). *Effective educational practices*. Educational practices series 3. Brüssel: International Academy of Education. Abgerufen von: <http://www.ibe.unesco.org>
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84(1), 30–43.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). What helps students learn? *Educational leadership*, 51(4), 74–79.
- Waxman, H. C., Wang, M. C., Anderson, K. A., & Walberg, H. J. (1985). Synthesis of research on the effects of adaptive education. *Educational Leadership*, 43(1), 26–29.
- Weber, T. (2002). *Kumulatives Lernen im Physikunterricht. Eine vergleichende Untersuchung in Unterrichtsgängen zur geometrischen Optik*. Dissertation an der Mathematisch-Naturwissenschaftlichen Fakultät an der Humboldt-Universität zu Berlin.
- Weinert, F. E. (1999). Aus Fehlern lernen und Fehler vermeiden. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlass des 60. Geburtstages von Fritz Oser* (S. 101–109). Opladen: Leske + Budrich.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Saganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Seattle, WA: Hogrefe & Huber Publishers.
- Weinert, F. E., & Helmke, A. (1996). Der gute Lehrer: Person, Funktion oder Fiktion? In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule* (34. Beiheft der Zeitschrift für Pädagogik, S. 223–233). Weinheim: Beltz.
- Weinert, F. E., & Helmke, A. (Hrsg.). (1997). *Entwicklung im Grundschulalter*. Weinheim: Psychologie Verlags Union.
- Weinert, F. E., Schrader, F.-W., & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research*, 13(8), 895–914.
- Weinert, F. E., Schrader, F.-W., & Helmke, A. (1990). Educational expertise. Closing the gap between educational research and classroom practice. *School Psychology International*, 11, 163–180.
- Wiater, W. (2008). *Unterrichtsprinzipien. Prüfungswissen – Basiswissen Schulpädagogik* (3. Aufl.). Donauwörth: Auer.
- Widodo, A., & Duit, R. (2004). Konstruktivistische Sichtweisen vom Lehren und Lernen und die Praxis des Physikunterrichts. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 233–255.
- Widodo, A., & Duit, R. (2005). Konstruktivistische Lehr-Lern-Sequenzen und die Praxis des Physikunterrichts. *Zeitschrift für Didaktik der Naturwissenschaften*, 11, 131–146.
- Wiens, P. D., Hessberg, K., Lo-Casale-Crouch, J., & DeCoster, J. (2013). Using a standardized video-based assessment in a university teacher education program to examine preservice teachers' knowledge related to effective teaching. *Teaching and Teacher Education*, 33, 24–33.
- Wiesner, H. (1995). Physikunterricht – an Schülervorstellungen und Lernschwierigkeiten orientiert. *Unterrichtswissenschaft*, 25(2), 127–145.
- Wiesner, H. (2006). Schülervorstellungen – eine vergangene Modeströmung in der Physikdidaktik? In R. Girwidz, M. Gläser-Zikuda, M. Laukenmann, & T. Rubitzko (Hrsg.), *Lernen im Physikunterricht* (S. 23–33). Hamburg: Kovač.



- Wiesner, H. (Hrsg.). (2008). Physikunterricht an Schülervorstellungen orientiert. *Praxis der Naturwissenschaften – Physik in der Schule*, 57(6).
- Wiesner, H., Schecker, H., & Hopf, M. (Hrsg.). (2011). *Physikdidaktik kompakt*. Freising: Aulis.
- Wilke, H.-J. (1993a). Zur Bedeutung des Experiments für den Physikunterricht. *Naturwissenschaften im Unterricht – Physik*, 4, 4–7.
- Wilke, H.-J. (1993b). Physikalische Schülerexperimente – Vorzüge, Erfahrungen und Probleme. *Naturwissenschaften im Unterricht – Physik*, 4, 8–11.
- Wilhelm, O., & Kunina-Habenicht, O. (2015). Pädagogisch-psychologische Diagnostik. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (2., Aufl., S. 305–328). Berlin: Springer.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Rating-skalen*. Göttingen: Hogrefe.
- Wolff, C. E., Bogert, N. van den, Jarodzka, H., & Boshuizen, H. (2015). Keeping an eye on Learning: Differences between expert and novice teachers' representations of classroom management events. *Journal of Teacher Education*, 66(1), 68–85.
- Zierer, K., Speck, K., & Moschner, B. (2013). *Methoden erziehungswissenschaftlicher Forschung*. München: Reinhardt.
- Zlatkin-Troitschanskaia, O., & Seidel, J. (2011). Kompetenz und ihre Erfassung – das neue „Theorie-Empirie-Problem“ der empirischen Bildungsforschung? In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Bildungsforschung. Traditionslinien und Perspektiven* (S. 218–233). Wiesbaden: VS Verlag für Sozialwissenschaften.





## Abbildungsverzeichnis

Abb. 1	Definition des Konstrukts Analysekompetenz und zeitliche Abfolge der beiden Validierungsstudien .....	15
Abb. 2	Das Kompetenzmodell von COACTIV mit Spezifikation für das Professionswissen (aus: Baumert & Kunter, 2011a, S. 32) .....	21
Abb. 3	Merkmale der Struktur- und Prozessqualität von Unterricht (aus: Plöger & Scholl, 2014c, S. 140) .....	24
Abb. 4	fünfstufiges Modell der Analysekompetenz (aus: Plöger et al., 2015, S. 170) .....	36
Abb. 5	Modell 2 und Modell 3 (aus: Plöger et al., 2015, S. 174) .....	42
Abb. 6	Modell 4 und Modell 5 (aus: Plöger et al., 2015, S. 175) .....	43

## Tabellenverzeichnis

Tab. 1	Zusammensetzung der Stichprobe nach Fachzugehörigkeit (in Anlehnung an Plöger & Scholl, 2014b, S. 280) .....	39
Tab. 2	Zuordnung der Items zur formalen und inhaltlichen Dimension (aus: Plöger et al., 2015, S. 178) .....	45
Tab. 3	Fit-Indizes der fünf Modelle (aus: Plöger et al., 2015, S. 180) .....	46
Tab. 4	Standardisierte Ladungen und Korrelationen auf latenter Ebene in Modell 5 (aus: Plöger et al., 2015, S. 180) .....	47
Tab. 5	Zusammensetzung des Samplings .....	58
Tab. 6	Beispiele für verwendete Transkriptionsregeln .....	61
Tab. 7	Zuordnung der deduktiv zu entwickelnden Kategorien zur Erfassung des Konstrukts Analysekompetenz .....	63
Tab. 8	Übersicht zu Fundstellen der jeweiligen Faktoren .....	68
Tab. 9	Ablaufmodell der Qualitativen Inhaltsanalyse .....	71
Tab. 10	Kategorien zur Überprüfung des Konstrukts Analysekompetenz .....	74
Tab. 11	Kategorie Zielorientierung .....	86
Tab. 12	Kategorie Strukturierung .....	86
Tab. 13	Kategorie Klarheit .....	87
Tab. 14	Kategorie Exemplarizität (Einheit von Konkretem und Abstraktem) .....	87
Tab. 15	Kategorie Lebensweltbezug des Themas .....	88
Tab. 16	Kategorie Verknüpfung mit dem Vorwissen .....	88
Tab. 17	Kategorie Aktivierung/Motivierung zur geistigen Selbsttätigkeit (kognitive Aktivierung) .....	89
Tab. 18	Kategorie Diskursivität des Unterrichtsprozesses (Reflexion von Inhalt und Methode) .....	90
Tab. 19	Kategorie Individualisierung/Differenzierung .....	90
Tab. 20	Kategorie Ergebnissicherung/Erfolgskontrolle .....	91
Tab. 21	Kategorie Anwendung/Transfer .....	91
Tab. 22	Kategorie Lernatmosphäre .....	92
Tab. 23	Kompetenzbereiche im Fach Physik (nach KMK, 2004a) .....	99
Tab. 24	Kategorie Physikalische Begriffsbildung .....	110

Tab. 25	Kategorie Durcharbeiten von Fachbegriffen .....	111
Tab. 26	Kategorie Experimentieren als zentrale Fachmethode .....	113
Tab. 27	Kategorie (weitere) fachliche Arbeitsweisen .....	114
Tab. 28	Kategorie Historische Genese physikalischen Wissens .....	116
Tab. 29	Kategorie Produktiver Umgang mit Fehlern .....	117
Tab. 30	Kategorien Lernen als konstruktive Tätigkeit von SchülerInnen .....	119
Tab. 31	Stufen des Kompetenzmodells von Plöger et al. (2015) in Entsprechung zu den Aspekten formaler Wissensqualität .....	122
Tab. 32	Kategorie Verstehenstiefe – Level 1 .....	126
Tab. 33	Kategorie Verstehenstiefe – Level 2a .....	127
Tab. 34	Kategorie Verstehenstiefe – Level 2b .....	128
Tab. 35	Kategorie Verstehenstiefe – Level 3a .....	129
Tab. 36	Kategorie Verstehenstiefe – Level 3b .....	130
Tab. 37	Überprüfung der Text-Kategorien-Zuordnung .....	136
Tab. 38	Anzahl der vorgenommenen Kodierungen für die inhaltliche und formale Dimension .....	137
Tab. 39	Anzahl der vorgenommenen Kodierungen für das didaktisch-methodische Wissen der ExpertInnen und NovizInnen .....	138
Tab. 40	Anzahl der vorgenommenen Kodierungen für das fachliche (incl. fachdidaktische) Wissen der ExpertInnen und NovizInnen .....	139
Tab. 41	Anzahl der vorgenommenen Kodierungen für Verstehenstiefe der ExpertInnen und NovizInnen .....	140
Tab. 42	Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für die inhaltliche und formale Dimension unter dem Aspekt des fachlichen Hintergrundes .....	141
Tab. 43	Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für das didaktisch-methodische Wissen unter dem Aspekt der Fachzugehörigkeit .....	143
Tab. 44	Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für das fachliche (incl. fachdidaktische) Wissen unter dem Aspekt der Fachzugehörigkeit .....	144
Tab. 45	Anzahl der vorgenommenen Kodierungen für ExpertInnen und NovizInnen für Verstehenstiefe unter dem Aspekt der Fachzugehörigkeit .....	145
Tab. 46	Überblick über die erhaltenen (vorläufigen) Reliabilitätswerte .....	159
Tab. 47	(vorläufige) Reliabilitäts- und Trennschärfewerte für die 19 Variablen der inhaltlichen Dimension (didaktisch-methodisches und fachliches Wissen) .....	160
Tab. 48	(vorläufige) Reliabilitäts- und Trennschärfewerte für die Teilskala didaktisch-methodisches Wissen (inhaltliche Dimension) .....	162
Tab. 49	(vorläufige) Reliabilitäts- und Trennschärfewerte für die Teilskala fachliches Wissen (FW) .....	163
Tab. 50	(vorläufige) Reliabilitäts- und Trennschärfewerte für die Teilskala Verstehenstiefe (VT) .....	164
Tab. 51	Endgültige Anzahl der vorgenommenen Kodierungen für die inhaltliche und formale Dimension .....	165
Tab. 52	Endgültige Reliabilitäts- und Trennschärfewerte für die Skala der inhaltlichen Dimension mit 16 Variablen .....	166

Tab. 53: Endgültige Reliabilitäts- und Trennschärfewerte für die Teilskala didaktisch-methodisches Wissen mit 10 Variablen .....	167
Tab. 54 Endgültige Reliabilitäts- und Trennschärfewerte für die Teilskala fachliches Wissen mit 6 Variablen .....	168
Tab. 55 Endgültige Reliabilitäts- und Trennschärfewerte für die Skala der formalen Dimension mit 5 Variablen .....	168
Tab. 56: Überblick über die endgültigen Reliabilitätswerte .....	169
Tab. 57 Korrelationen der Summen der Kodierereignisse der drei Teilskalen untereinander und mit der Gesamtsumme aller Kodierereignisse .....	170
Tab. 58 Mittelwerte, Standardabweichungen und Standardfehler für die ermittelten Kodierereignisse nach Gruppen und Teilskalen .....	170
Tab. 59 Kolmogorow-Smirnow-Test auf Normalverteilung .....	171
Tab. 60 Mann-Whitney-Test (U-Test) .....	172
Tab. 61 Effektstärken .....	172
Tab. 62 Transkriptionsregeln .....	216
Tab. 63 Trainingsprogramm für die KodiererInnen-Schulung .....	219



# Anhang

## Anhang A: Interviewleitfaden

---

### Begrüßung durch den Interviewer

---

Vielen Dank, dass Sie sich Zeit genommen haben, an dieser Studie teilzunehmen. Ich erkläre Ihnen das weitere Vorgehen: Wir werden uns nun Ausschnitte aus einer Physik-Stunde ansehen. Es handelt sich um eine 45-minütige Unterrichtsstunde, die auf 15 Minuten zugeschnitten wurde, ohne dass der Prozesscharakter der Stunde verloren geht. Die Stunde wurde zu Demonstrationszwecken aufgenommen. Es handelt sich also um keine ganz natürliche Situation. Die Authentizität ist aber dennoch hoch. Ich werde die Aufnahme zwischendurch 4mal anhalten und Ihnen Gelegenheit geben, das Beobachtete zu kommentieren. Haben Sie noch Fragen zur Vorgehensweise?“ (Papier und Stift bereitstellen)

---

### Start Video zu Abschnitt I, Stopp bei 3:40 Minuten

---

„Hier machen wir den ersten Cut. Kommentieren Sie doch bitte, was Sie im ersten Abschnitt dieser Unterrichtsstunde beobachten konnten.“

„Wie führt der Lehrer das Problem der Lichtbrechung ein?“

„Was sagen Sie zu der Auswahl der Beispiele?“

„Haben Sie noch weitere Anmerkungen zum ersten Abschnitt?“

„Jetzt sehen wir uns den weiteren Verlauf der Stunde an.“

---

### Start Video zu Abschnitt II, Stopp bei 8:20 Minuten

---

„Hier machen wir den zweiten Schnitt. Nehmen Sie bitte auch zu diesem Abschnitt Stellung.“

„Was sagen Sie zur Weiterentwicklung des Stundenthemas?“

„Gibt es noch weitere Unterrichtsmomente, zu denen Sie Stellung nehmen möchten?“

„Dann sehen wir uns nun den weiteren Verlauf der Stunde an, in der Zwischenzeit ist das Experiment in Partnerarbeit durchgeführt worden.“

---

### Start Video zu Abschnitt III, Stopp bei 10:46 Minuten

---

„Nehmen Sie bitte auch zum dritten Ausschnitt der Stunde Stellung.“

„Wie beurteilen Sie die Vorgehensweise bei der Auswertung der Messergebnisse?“

„Möchten Sie noch ergänzende Anmerkungen vornehmen?“

„Jetzt folgt ein sehr kurzer Abschnitt. Inzwischen ist das Brechungsgesetz theoretisch erarbeitet worden, was durch den Videoschnitt für uns nicht sichtbar war.“

---

### Start Video zu Abschnitt IV, Stopp bei 12:10 Minuten

---

„Was haben Sie in diesem Unterrichtsausschnitt beobachten können?“

„Nehmen Sie bitte auch zu diesem Ausschnitt der Stunde Stellung.“

„Möchten Sie weitere Ausführungen machen?“

---

### Start Video zu Abschnitt V, Stopp am Ende des Videos

---

„Was haben Sie im letzten Teil der Unterrichtsstunde beobachten können?“

„Wenn wir jetzt auf den gesamten Unterrichtsablauf zurückblicken: Welche Aspekte dieser Unterrichtsstunde erscheinen Ihrer Ansicht nach positiv und/oder negativ hervorhebenswert?“

Frage nach Alternativen: „Würden Sie das Thema in ähnlicher Weise behandelt haben oder sehen Sie Alternativen?“

„Abschließend würde ich Sie bitten, sich zu überlegen, ob es noch den einen oder anderen Moment gibt, den Sie kommentieren möchten.“

„Vielen Dank für das Gespräch und ihre Einschätzungen.“

---

## Anhang B: Transkriptionsregeln – Teil 1 (allgemeine Hinweise)

Bitte bei der Transkription auf folgende Punkte achten:

- Die interviewten Personen werden im Transkript abgekürzt und anonymisiert dargestellt.
- Der Interviewtext wird in Standardorthografie aufgeschrieben.
- Es erfolgt keine Großschreibung nach Satzzeichen, denn Satzzeichen sollen die Intonation anzeigen.
- Hauptwörter werden im Transkript großgeschrieben.
- Beim Neuansetzen eines Sprechers wird das erste Wort mit Großbuchstaben begonnen.
- Die Kommasetzung erfolgt im Sinne bester Lesbarkeit nach den Regeln der deutschen Rechtschreibung.
- Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend.
- Es werden wiederholte Wörter, abgebrochene Wörter und Sätze notiert.
- Dialektale Ausdrücke, die besonderes (regionales) Kolorit besitzen bzw. schwer übersetzbar sind, werden mit übernommen.
- Pausen werden durch (.) oder entsprechende Zeitangabe [z.B. (2); (3)] markiert.
- Betonungen in der Aussprache der interviewten Personen werden hervorgehoben.
- Zustimmung oder bestätigende Lautäußerungen der Interviewer oder interviewten Personen (z.B. hmm) werden nicht extra als Redebeitrag markiert, sondern in den Redefluss der anderen Person integriert, z.B. durch //hmm//.
- Das Ende einer Aussage wird durch eine Zeitangabe markiert, z.B. #00:07:30-3#.



**Anhang C: Transkriptionsregeln – Teil 2****Tab. 62:** Transkriptionsregeln

<b>Zeichen</b>	<b>Bedeutung</b>
.	deutlich fallende Intonation
;	leicht fallende Intonation
?	deutliche Frageintonation
,	leicht steigende Intonation, die Anschlussäußerung erwarten lässt
(.)	kurze Sprechpause oder Verzögerung
(2)	lange Sprechpause mit Sekundenangabe
brau-	Wortabbruch
<hh>	hörbares Einatmen (Anzahl der h entspricht Länge der Atmung)
Wort	Betonung
°Wort°	gesenkte Lautstärke/undeutliches Sprechen
Nein	Laut in Relation zur üblichen Lautstärke des Sprechers
@nein@	lachend gesprochene Worte
ja::	Dehnung (Doppelpunktzahl entspricht Länge der Dehnung)
nich'	Buchstabenausfall
so=ne	Wortverschmelzung; Wortverschleifung
(vielleicht)	unsichere Transkription
(...)	unverständliche Transkription
((räuspert sich))	Kommentar
[	Beginn eines Beitrags mit unmittelbarem Anschluss oder Überlappung
//mh//	Äußerung des Interviewpartners im Hintergrund
gut richtig	gleichzeitig gesprochene Beiträge

## Anhang D: Auszug aus einem Interviewtranskript

### Transkriptionskopf

ProbandIn: Exp-NW 1

Interviewer: Prof. Dr. W. Plöger, Uni Köln

Datum: 14.03.2014

Dauer des Interviews: 38 Minuten

Transkription: Max Peltzer

Korrektur: Eric Berndt

2. Korrektur: Alina Diederichs

**IW-P:** Hier machen wir den ersten Cut, ((räuspert sich)) kommentieren Sie doch bitte, was sie im ersten Abschnitt der Unterrichtsstunde beobachten konnten, #00:00:02-7#

**Exp-NW1:** Ja ähm der Lehrer hat das Thema angegeben worum es gehen soll, <h> und das ist offensichtlich die Einstiegsphase in die Stunde //hmhm// also er knüpft dann nochmal an an 'ne voraus- ä vorangehende Stunde, <h> ist vielleicht die zweite Hälfte einer Doppelstunde ich weiß es nicht. <hh> ähm (.) methodisch gibt er nich' klar an wie es laufen soll //hmhm// also was jetzt äh geschieht, <h> er äh es is' auch erstmal ni- unklar ob er jetzt den deduktiven oder induktiven //hmhm// Ansatz hat. also ob er er nennt das Gesetz Brechungsgesetz <h> ob er erst das Gesetz vorstellen will, das könnte man ja machen //hmhm// Gesetz vorstellen und dann <h> durch Experimente, Beobachtungen sonst was das erarbeiten oder erklären, <h> oder ob er anders vorgehen will ob er äh 'ne Beobachtung äh ermöglichen will oder 'n Experiment ähm aufstellen. <hh> ähm das bleibt letztlich unklar eigentlich in der ganzen Phase //hmhm// äh das ist recht unsortiert bei ihm //also Beobachtung äh ausgehend von der Beobachtung dann äh das Gesetz induktiv sozusagen zu erarbeiten? das meinen sie// genau //ja, hmhm// er könnte ja auch ohne weitere Erläuterung einfach ein Phänomen vorstellen //ja, hmhm// das beobachten lassen //hmhm// deuten lassen und von da aus dann zu dem Gesetz kommen. //genau// also die Schüler das selber verallgemeinern lassen, <h> macht er aber nich' er unterbricht sich auch selber in seinem äh Lehrervortrag, //hmhm// mehrmals kommt wir sind in 'ner Situation 400 Jahre zurück, <h> ähm dann spricht er auch meine ich an zwei Stellen davon wir wollen jetzt ein Experiment machen, <h> es kommt aber doch nicht, ä zwischendurch kommt dann=ne <h> Bildpräsentation, <h> ähm er sagt wir müssen uns fragen an beiden, an mehreren Stellen <h> ähm aber die Fragen stellt er auch dann selber //hmhm// also insgesamt der Redeanteil is' gefühlte 99 Prozent //ja// würd' ich sagen, es gibt zwei Schülerantworten, //hmhm// wovon er die erste <h> direkt kommentiert, die zweite so etwas ironisch äh konnotiert, //hmhm// als eine Schülerin <h> äh was über=n Lichtstrahl im Wasser sagt geht=s da um=s Fischen. //hmhm// <hh> ähm (.) ja, also er (.) nochmal er f- gibt keinen klaren methodischen Weg an wie er vorgehen möchte, <h> und die Wege die er so ähm wo er ansetzt die unterbricht er selber auch er bleibt sich da nicht treu, <h> ähm zum Beispiel präsentiert er ja dann Bilder die hätte man ja erstmal genau beschreiben und dann analysieren //hmhm// können, macht er aber auch nicht, <h> sondern er er beschreibt das alles selber, auch immer wieder auf so=ner ganz theoretischen Ebene, <h> und fachsprachlich aber auch nicht korrekt. also äh da kommen sch- mehrmals so ähm Formulierungen wie da ist auch Lichtbrechung im Spiel //hmhm// was heißt das //hmhm// also entweder wird

da das Licht //ja// gebrochen oder nicht, das kann man dann ja auch klar benennen. <h> ähm (.) ja also an mehreren Stellen ist das auch äh ja fachlich unsauber <h> //hmhm// zuerst sagt er dann wir wollen uns angucken Lichtstrahl im See, aber dann sagt er naja 'n See ist es nicht sondern 'n Eimer. <h> ähm also das kann man ja alles machen Modellexperimente und so weiter //ja// aber dann muss man es klar so benennen. #00:03:25-2#

**IW-P:** Hmhm (.) ja <h> also das ist ja jetzt eine Einleitungsphase wie sie ja selbst gesagt haben. äh für mich ist nochmal die Frage ob sich aus dem was bisher gesagt wurde und vor allen Dingen was auch die Bilder die demonstriert wurden anbetrifft, <h> sich für die Schülerinnen und Schüler überhaupt eine Problemstellung eine Fragehaltung äh ergibt?

**Exp-NW1:** [ Ne es ist überhaupt kein Problem benannt, also das Gesetz ist mal benannt worden <h> und eben in so diffusen Formulierungen dass alles //ja, ja// alles hat mit allem was zu tun, da ist Licht überall im Spiel. //ja, ja// also zum Beispiel äh diese Fotos die dann äh dieses Flirren da auf der Straße und so weiter zeigen das ist ja nochmal physikalisch, denke ich mal, sind das ganze andere Vorgänge als //so ist es// jetzt ä der Lichtstrahl im Wasser. #00:04:30-4#

**IW-P:** [ Ja das ist also eigentlich überhaupt kein treffendes Beispiel dafür, ähm das das ist viel zu komplex, (und) hätte er eigentlich fortlassen müssen genau. (.) ja sollen wir=n Stück weiter gehen? //hmhm// #00:04:42-4#

**VA:** #00:04:55-6#

## Anhang E: Trainingsprogramm für die KodiererInnen-Schulung

**Tab. 63:** Trainingsprogramm für die KodiererInnen-Schulung

	Inhalt(e)	Ziel(e)
<b>Tag 1</b>	<p>Einführung in das Forschungsprojekt</p> <p>Inhaltliche Schwerpunkte:</p> <p>theoretische Einführung in die Fragestellung der Untersuchung</p> <p>Konzeption des Konstrukts „Analysekompetenz“</p> <p>Vorstellung des methodischen Vorgehens der Untersuchung</p> <p>Vorführen des Videos (mehrmaliges Anschauen)</p> <p>Einführung in das Kategoriensystem (Präsentation der zugrunde liegenden Dimensionen und Faktoren)</p> <p>Besprechung der einzelnen Kategorien mit ihren Definitionen, Beschreibungen und Ankerbeispielen</p>	<p>gegenseitiges Kennenlernen des KodiererInnen-Teams;</p> <p>Vorstellung und Einarbeitung in das Forschungsprojekt;</p> <p>Vertrautmachen mit den inhaltlichen Details der Unterrichtsstunde;</p> <p>Vertrautmachen mit dem Kategoriensystem;</p> <p>Einarbeitung in das Kategoriensystem</p>
<b>Tag 2</b>	<p>Einstieg: Wiederholung des Kategoriensystems; Klären von Fragen</p> <p>Übung an zwei Transkripten</p> <p>gemeinsame Kodierung in drei Durchläufen:</p> <p>Durchlauf 1: Anwendung der Kategorien zum didaktisch-methodischen Wissen</p> <p>Durchlauf 2: Anwendung der Kategorien zum fachlichen (incl. fachdidaktischen) Wissen</p> <p>Durchlauf 3: Anwendung der Kategorien zu den Levels der Verstehenstiefe</p> <p>sofortige Rückmeldung und Korrektur der Kodierungen</p>	<p>Kategorien (incl. Definitionen und Beschreibungen) in Erinnerung rufen;</p> <p>Sicherheit im Umgang mit den Kategorien gewinnen;</p> <p>Fehler besprechen; Fehlkodierungen tilgen</p>
<b>Tag 3</b>	<p>Einstieg: Wiederholung des Kategoriensystems, der Kategorien und Besprechung der Kodierungen (insb. Klärung von offenen Fragen)</p> <p>Übung an zwei Transkripten</p> <p>selbständige Kodierung in drei Durchläufen:</p> <p>Durchlauf 1: Kategorien zum didaktisch-methodischen Wissen</p> <p>Durchlauf 2: Kategorien zum fachlichen (incl. fachdidaktischen) Wissen</p> <p>Durchlauf 3: Kategorien zu den Levels der Verstehenstiefe</p> <p>Korrektur der Kodierungen der KodiererInnen; Besprechen von Fehlkodierungen und „kritischen“ (überarbeitungswürdigen) Stellen bei der Kategorienformulierung</p>	<p>Sicherheit im Umgang mit den Kategorien gewinnen;</p> <p>Aufbau der entsprechenden Interpretationskompetenz;</p> <p>routinierter Umgang mit den Kategorien;</p> <p>Vertiefung und Verbesserung der Kodierfähigkeiten;</p> <p>Sensibilisierung für das Kodierungsverfahren</p>



**Die vorliegende qualitative Studie „Wie analysieren ExpertInnen und NovizInnen Unterricht?“ untersucht die Analysekompetenz von Lehrpersonen – verstanden als jene Fähigkeit, die Qualität von Unterricht unter dem Aspekt der Lernwirksamkeit angemessen beurteilen zu können.** Im Fokus stand dabei die Frage, welche Faktoren die Ausprägung dieser Kompetenz bestimmen. Die notwendigen Daten wurden durch leitfadengestützte Interviews generiert, in denen jeweils neun ExpertInnen und NovizInnen eine videografierte Unterrichtsstunde analysierten. Die inhaltsanalytische Auswertung der auf diese Weise gesammelten Aussagen zeigte, dass die ExpertInnen – im Gegensatz zu den NovizInnen – über umfangreiches pädagogisches sowie fachliches (incl. fachdidaktisches) Wissen verfügten und mit diesem Wissen nicht nur einzelne Situationen, sondern auch den Unterrichtsprozess als komplexes Ganzes erfassen konnten.



#### **Der Autor**

**Matthias Krepf**, Jahrgang 1985, promovierte 2018 an der Humanwissenschaftlichen Fakultät der Universität zu Köln. Im Anschluss an sein Lehramtsstudium arbeitete er an der Universität zu Köln zunächst am Lehrstuhl für Allgemei-

ne Didaktik und Schulpädagogik und ist seit 2018 als Wissenschaftlicher Mitarbeiter und Lehrkraft für besondere Aufgaben am Lehrstuhl für Empirische Schulforschung (Schwerpunkt quantitative Methoden) angestellt. Seine Forschungsinteressen liegen in der empirischen Lehrer\*innenbildungsforschung sowie der Schul- und Unterrichtsforschung.

978-3-7815-2312-8

